DEPARTMENT OF DEFENCE
DEFENCE SCIENCE & TECHNOLOGY ORGANISATION
DSTO

Representing Uncertainties Using Bayesian Networks

Balaram Das

DSTO-TR-0918

Representing Uncertainties Using Bayesian Networks

Balaram Das

Information Technology Division
Electronics and Surveillance Research Laboratory

DSTO-TR-0918

ABSTRACT

This report demonstrates the application of Bayesian networks for modelling and reasoning about uncertainties. A scenario for naval anti-surface warfare is constructed and Bayesian networks are used to represent and update uncertainties encountered in the process of 'situation assessment'. Concepts from information theory are used to provide a measure of uncertainty and understand its flow in a Bayesian network. This in turn yields analytical methods to formulate various effectiveness measures.

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE & TECHNOLOGY ORGANISATION
DSTO

Published by

DSTO Electronics and Surveillance Research Laboratory
PO Box 1500
Salisbury South Australia 5108 Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-011-177
December 1999

APPROVED FOR PUBLIC RELEASE

Representing Uncertainties Using Bayesian Networks

Executive Summary

The work reported here was undertaken in relation to a broader task which is aimed at providing better tools and techniques in aid of command, control, communications and intelligence (C3I). The task plan places major emphasis on using the techniques of modelling and simulation in the analysis and resolution of C3I problems. A crucial problem that decision makers face in any C3I process is the problem of uncertainty. Here we use Bayesian networks to model uncertainty and reason about it in both a qualitative and a quantitative manner. It is hoped that the tools developed here would be integrated with other existing simulation tools to provide a refined and more versatile modelling environment and decision aid package.

The central features of the Bayesian network approach are:

- *Qualitative:* Given a scenario, a Bayesian network depicts graphically the cause and effect relationship between various elements of the scenario. In doing so it also demonstrates conditional independence i.e., which factors are relevant and directly affect a given event and which factors are irrelevant – irrelevant in the sense that knowledge regarding these factors become redundant once the direct causes are known.
- *Quantitative:* It updates probability distributions. Given a battlefield *situation* and a prior probability distribution over a hypothesis variable that represents possible enemy courses of action, Bayesian network provides the capability to update this probability distribution when fresh reconnaissance and surveillance data are obtained.

The pictorial display of the model as a graph facilitates easy understanding and is therefore of great help in rapid model development.

To make the analysis specific we construct a scenario of naval anti-surface warfare. The scenario has the advantage of being based upon an approved naval exercise called *Operation Dardanelles*. Bayesian networks are constructed to represent the uncertainties in the process of *situation assessment*. A temporal development of the scenario is considered, which evolves through a number of stages starting from the initial detection of the enemy to a major engagement between Blue and Orange forces. Bayesian networks are used at every stage to update knowledge and decide upon a *course of action* (COA). These networks have been implemented through the commercially available product *Netica*.

After demonstrating the applicability of Bayesian networks to command and control problems, we turn our attention towards developing analytical tools to investigate the flow of information and uncertainty in the network. Concepts from information theory are used to provide a measure of uncertainty, a measure of sensor effectiveness, a measure of the effectiveness of belief updating and finally a measure of the effectiveness of any particular Bayesian network considered as a decision aid tool. A number of examples related to the scenario are considered and numerical effectiveness measures obtained.

Author

Balaram Das
Information Technology Division

Balaram Das works for the System Simulation and Assessment Group. His work mainly deals with the application of mathematical methods to problems pertaining to command and control. His current work utilises stochastic methods to model uncertainties.

Contents

1. Introduction

The work reported here was undertaken in relation to a broader task which is aimed at providing better tools and techniques in aid of command, control, communications and intelligence (C3I). The task plan places major emphasis on using the techniques of modelling and simulation in the analysis and resolution of C3I problems and major simulation tools have been developed under this task. A crucial problem that decision makers face in any C3I process is the problem of uncertainty. In fact many decision makers contribute to the view that good decision making is all about handling uncertainties properly. This is probably an extreme point of view but it emphasises the ever present effects of uncertainty on all aspects of decision making. This report documents an effort to find appropriate tools and techniques for constructing *probabilistic models* to represent and analyse uncertainties encountered in the process of command and control. The aim is to build models and develop analytical methods for both qualitative and quantitative analysis of uncertainty. It is hoped that the tools developed here would be integrated with other existing simulation tools to provide a refined and more versatile modelling environment and decision aid package.

Of course there exist other methods, often mentioned in artificial intelligence and expert systems literature, that also handle uncertainty quite adequately, viz. fuzzy logic, belief functions etc., to mention a few [Kanal and Lemmer 1986, 1988; Dubois, Prade and Yager 1997]. However, a probabilistic approach has the advantage that it is based on a rigorous theory with a vast amount of known results. This is a great advantage and has in fact prompted many to claim that probability is the only sensible description of uncertainty and is adequate for all purposes [Lindley, 1987; Cheeseman, 1986]. The opposing school of researchers point out that probability requires the enumeration of all possibilities, which in turn requires a vast amount of storage and computational manipulation making probabilistic methods computationally infeasible. It is to overcome this computational impasse that Bayesian networks were formulated [Pearl 1988, Neapolitan 1990]. This work adopts the Bayesian interpretation of uncertainty and uses Bayesian networks to build computational models. The emphasis is on obtaining numerical measures, wherever possible, of the effects we are interested in.

Bayesian networks have been used extensively to model real world problems [Ottonello et al., 1992; Burnell and Horvitz, 1995; Fung and Del Favero, 1995; Heckerman, Breese and Rommelse, 1995] in particular it has been used very successfully in building expert systems to help medical diagnosis [Spiegelhalter et al., 1993]. Levitt and his coworkers [Levitt et al. 1995] have proposed a formalism, that uses Bayesian networks to analyse synthetic aperture radar imagery to probabilistically rank interpretations of the presence and location of military forces on the ground. Chang et.al. [Chang et. al. 1996; Chang and Fung 1997; Lui and Chang 1996] have considered a number of problems in target recognition to which Bayesian networks can be fruitfully applied. Staker [Staker 1999] uses Bayesian Networks to assist

commanders in determining the level of risk associated with their information system networks and Manka and Nicholson [Manka and Nicholson, 1999] discuss how Bayesian networks can help special forces to reason about enemy intent. Owing to its popularity a number of software packages are available to implement and manipulate Bayesian networks [Almond 1996]. In this respect Fabian and Lambert argue that traditional Bayesian networks offer a limited capacity for knowledge representation [Fabian and Lambert 1998] and offer a solution which has been implemented in the ATTITUDE architecture [Lambert 1998]. The networks constructed in this report have been implemented through the commercially available product *Netica* [Netica 1998].

We start by introducing the concept of probabilistic models and in appendix B we discuss Bayesian networks with just enough details to make this document self contained.

In sections 2 we analyse the problem of *'situation assessment'* [Endsley 1995]. Bayesian networks are constructed to represent the uncertainties in this process. To make the analysis specific we construct a scenario of naval anti-surface warfare. The scenario has the advantage of being based upon an approved naval exercise called *'Operation Dardanelles'*. A temporal development of the scenario is considered, which evolves through a number of stages starting from the initial detection of the enemy to a major engagement between Blue and Orange forces. Bayesian networks are used at every stage to update knowledge and decide upon a *course of action* (COA).

After this particular demonstration of the applicability of Bayesian networks to command and control problems, we turn our attention towards developing analytical methods to investigate the flow of information and uncertainty in the network. In section 3 information theoretical concepts are used to provide a measure of uncertainty, a measure of sensor effectiveness, a measure of the effectiveness of belief updating and finally a measure of the effectiveness of any particular Bayesian network considered as a decision aid tool. A number of examples related to the scenario are considered and numerical effectiveness measures obtained.

Finally section 4 concludes this report with suggestions for future research.

1.1 Probabilistic Models

Models which capture uncertainties in terms of probabilities would be very relevant to our endeavour. For, probabilities not only provide us with numerical estimates to weigh alternatives against each other, but the calculus of probability also provides us with a rigorous procedure to articulate and manipulate qualitative relationships amongst alternatives. In our analysis here we will almost always adopt the *subjective interpretation of probability* [Appendix A]. This is because, when decision makers dealing with C3I problems quantify the likelihood of events, they generally base their estimates on their personal knowledge. They take into account the relevant information available, their past experience and sometimes even their intuition and prejudices to

decide upon a probability factor. This leads to the Bayesian formalism, the centre piece of which is the Baye's rule, that allows a decision maker to update his subjective belief when new facts are uncovered [Appendix A].

To clarify the basic notations; we denote random variables by capital letters X, Y, Z, ..., etc.- or by subscripted letters X_1, X_2, X_3, ..., etc. The values taken by these variables are denoted by lower case letters x_i, y_j, z_k, ..., etc.- or correspondingly by x_{1i}, x_{2j}, x_{3k}, ..., etc. The probability for X to assume the value x_i is denoted by $P(X=x_i)$ or in short $P(x_i)$.

Suppose we have built a model of some C3I problem and the uncertainties in the model have been captured by, say, the four discrete random variables, W, X, Y and Z. The model will be an adequate model if it encodes all probabilistic information that permit us to calculate all marginal, conditional and joint probabilities like say, $P(x_i)$ or $P(x_i \mid y_j)$ etc. To be able to do this the model must provide the values of all joint distributions $P(w_i, x_j, y_k, z_l)$. In fact this is the main aim of probabilistic modelling: *once a problem domain has been modelled through random variables, to provide a joint distribution over these variables that captures the uncertainties inherent in the system being modelled.* In other words by a *probabilistic model* we mean [Pearl 1988]:

> A set $\boldsymbol{U}$ of discrete random variables together with a joint probability distribution $P(\cdot)$ defined over these variables.

1.1.1 Computational Feasibility of Probabilistic Modelling

If the random variables W, X, Y, Z, mentioned above are simple propositional variables we would require a storage with 2^4 places. If our problem is modelled by n random variables we would require at least 2^n places of storage capacity, i.e. a storage capacity that grows exponentially. What is more, if we want to compute the conditional probability $P(x_i \mid y_j)$ say, then the rules of probability theory assert that we must evaluate the following

$$P(x_i \mid y_j) = \frac{P(x_i, y_j)}{P(y_j)} = \frac{\sum_{kl} P(w_k, x_i, y_j, z_l)}{\sum_{kil} P(w_k, x_i, y_j, z_l)}$$

This entails dividing two marginal probabilities each involving a summation over an exponentially large number of variable combinations. In other words, we not only require a large storage capacity, we also require a large number of computational steps to obtain the information we need from the model. This computational impasse was overcome when it was realised that in most estimates we do not necessarily need the complete array of all values of joint distribution. When estimating the probability of some situation we only need to know the context dependent probabilities for that situation. In other words, we only need to know the conditional probabilities of the situation in question with respect to the causes that directly influence the situation. Fortunately probability theory has a unique ability to process context sensitive beliefs. Furthermore, context dependency or more to the point context independency can be

represented by graphs and manipulated by local propagation [Pearl 1988]. This is precisely what Bayesian networks do thereby making probabilistic modelling computationally feasible.

Given any C3I problem we start by building a Bayesian network that represents the uncertainties in that problem. The required probabilistic model then emerges from the network. We explain this in a easy to understand fashion in Appendix B. Readers familiar with Bayesian network may proceed directly to the next section.

2. Situation Assessment with Bayesian Networks

The art of decision making under uncertainty is essentially the art of situation assessment when data are lacking [Discenza]. If the true situation can be ascertained, deciding upon an action would essentially be mechanical. In its bare essentials, situation assessment consists of integrating information from different parts and different elements of the battlefield to form the current total picture, and predict future trends. The challenge of making global assessments from local information is of course the most difficult part of the process. There are few commonly agreed upon procedures for such integration, different experts adopt different methods depending upon their experience and personal reasoning process. Situation assessment therefore contains a large component of subjective judgement. As Bayesian networks are particularly good at capturing subjective judgements, they provide a unique tool for modelling this process.

Chang [Chang 94] considers the problem of multiple intelligence correlation and fusion using Bayesian network in order to identify enemy targets and infer about their mission. He acquires domain knowledge for the ground tactical scenario with ELINT and COMINT sensors, and the constructed Bayesian network models the characteristics of the intelligence data. A Bayesian network inference algorithm is provided. Results for a test scenario are also documented to illustrate the data fusion. Discenza [Discenza] combines reconnaissance information, both from direct observations and failures to detect, regarding individual units in the battlefield. Bayes rule supported by Monte-Carlo motion models is used for updating individual unit's position and its probable track. These Bayesian updates are then used to determine *a posteriori* probabilities for alternate enemy courses of action.

2.1 Multi-Stage Decision Making

We view decision making as essentially a multi stage process. Any stage of decision making starts with the input which are observational data collected from diverse sources. These information sources can be a series of sensors using electromagnetic or sound waves. They can also be direct observations of the enemy obtained through

reconnaissance missions. The information may include; reports of target detection, its mobility and its activities like communication, EW, or weapon launch etc.

The gathered information is then fused to minimise noise in the information and to get the best possible appreciation of the physical and behavioural aspects of the enemy. Essentially these would involve, estimating location, tracking and discrimination between various enemy elements, and finding out the activities they are engaged in.

With the state of the enemy thus appraised, an assessment of the situation is next conducted, which among other aspects involves assessing enemy intention, the level of threat it poses, and predicting its future behaviour. The decision maker is now in a position to weigh the appropriateness of alternate courses of action and decide upon one - and the cycle starts again [Wohl 1981; Waltz and Buede 1986].

2.2 Generating Hypothesis

As indicated above, the multistage decision process starts with the arrival of initial stimuli in the form of intelligence reports about enemy activity. A set of hypotheses are then formed to make out what the enemy might do - its intention. Assessing *enemy intention* involves understanding enemy plan of action, and the knowledge requirement for anticipating enemy plans is tremendous. Even when one has an accumulation of sufficient prior intelligence, the process of representing this knowledge and reasoning about them is rather difficult. In fact, this will involve having at hand a database that records patterns of enemy past behaviour, plus an analysis of enemy warfare doctrine, and an understanding of various operational procedures. One should then be able to deduce a pattern from the current set of activities, and match the current pattern with past patterns of activity and operational procedures. The problem of hypothesis generation and assigning prior probabilities to the hypotheses are considered with respect to a given scenario by Laskey et al. [Laskey 1994]. They then use Bayesian networks to construct probabilistic models to assist an intelligence analyst in evaluating how well various hypotheses about the adversary's intentions are supported by available evidence. For the purpose of this analysis we assume that the hypotheses generation has been done externally and a set of initial hypotheses are available for starting the situation assessment process.

Hypotheses representing enemy intent are not directly observable entities. Therefore, a set of *information variables* have to be devised which can be observed and which will provide sufficient details to assign likelihoods to the set of hypotheses. The commander then directs the sensors and other intelligence gathering agencies to gather information regarding these information variables.

2.2.1 Incompleteness of Hypothesis Set

An important problem relates to the fact that it is impossible to generate a complete set of hypotheses, a set that exhausts all enemy options. For this reason we add the *'not modelled'* hypothesis to our set of hypotheses. This allows for the possibility that the true enemy intent has not been fathomed and in general this hypothesis should be endowed with a low value of probability at the outset. If during the course of situation assessment the posterior probability of the *not modelled* hypothesis grows large then we infer that "we do not know" what the enemy is doing. A new set of hypotheses must be generated and the analysis reworked.

2.3 Constructing a network

With the hypothesis and information variables thus formulated a Bayesian network can now be constructed. The root node of such a network would contain the hypothesis variable whose states corresponds to possible enemy intent. The information variables occupy the lowest level nodes; nodes that do not have any children. In general a network will have a number of intermediate nodes. The hypothesis node is causally linked to the information nodes through these intermediate nodes. The nodes and the links should reflect the causal structure and context independencies pertaining to the situation assessment task at hand. These concepts are explained in appendix B and will become clear when we construct illustrative networks below. Evidence regarding information variables is gathered through sensors. The information from this evidence propagates against the links of the network to update the probability distribution over the hypothesis variable. The causal structure of the network therefore encapsulates the reasoning process that the user employs to reason about the likelihood of any hypothesis state given the current nature of evidence.

2.3.1 Generating Conditional Probabilities

The bunch of parent to child links in the network contain the conditional probabilities $P(c \mid p_1,p_2, \ldots ,p_n)$ – this gives the probability of observing the child in state c, given that the n parents are in states $p_1, p_2, \ldots ,p_n$ respectively. How would one generate these conditional probabilities? We stipulate that these be gleaned from the commander's subjective knowledge, by some process of querying his or her beliefs. Essentially we are advocating that the commander's *cognitive process,* rather than an algorithmic *rule based process* which form the basis of all expert systems, be used to produce the conditional probability table. The commanders are very capable experts with considerable experience. While forming a judgement they can pick and bring together the pertinent facts along with the past experiences that are of relevance. There are no difficulties in creating a database that can store all the necessary facts, together with a rule base that encapsulates the collective wisdom of various commanders. But there may not exist an algorithm that is capable of extracting what is needed from the morass of data

and producing the appropriate inference. Even when one is found it may not be practical. In fact according to Moffat [Moffat 98] such an approach often produces very complex models which are difficult to understand and slow in running.

The problem of modelling a commander's intuitive knowledge and generating useful information from this model to enable one to construct a network is a difficult problem. A starting point for research in this direction can be the *naturalistic decision making models* proposed by Klien and others [Zsambok and Klien 1997]. Here we assume that the commanders subjective knowledge has been quantified to generate the conditional probability tables required by the network.

2.4 Input into a Network

A Bayesian network accepts evidence in two distinct categories:

- Evidence that sets the prior probability distribution over the hypothesis variables. Such evidence is therefore injected through the hypothesis variable and travels along the links of the network to set the prior probabilities of all other variable.
- Evidence regarding the information variable. These are injected through the information variables and travel against the links to update the prior probabilities to posterior probabilities according to Bayes rule.

2.4.1 Generating Prior Probabilities

Every stage of situation assessment requires assigning prior probabilities to the hypotheses. These prior probabilities are obtained from a knowledge of the prevailing situation. This problem of converting a *state of knowledge* to a probability assignment is a problem that lies at the heart of Bayesian probability theory. When the prior information can be given a precise mathematical formulation it is often possible to prescribe a set of rules for converting the information to a probability assignment. The general problem, however, is far from being solved and continues to be a field for continuing research [Bernardo and Smith 1994]. We shall assume that there exist external modules which convert the current state of knowledge to a prior probability distribution over the hypothesis variable.

2.4.2 Evidence for Updating to Posterior Probabilities

As discussed above, evidence pertaining to information variables is used for belief updating. Such evidence is gathered through surveillance and reconnaissance. It is always preferable to use more than one sensor in the information gathering process, for there are fundamental limitations on any attempt at situation assessment based on the observations of a single sensor [Durrant-Whyte 1991]:

- Single sensors can only provide partial information about the situation being monitored.
- There is no way of reducing noise related uncertainty or testing observations for errors.
- Different sensors can provide qualitatively different observations which would be useful to tackle different situations, no single sensor can provide observations that can cover all eventualities.
- Single sensor systems are not robust as a failure of the sensor will result in complete system failure.

It is therefore an accepted paradigm in military command and control situations to integrate information obtained from a number of geometrically, geographically and physically diverse sensors to overcome the limitations inherent in the use of single sensors. The potential advantages of integrating information from multiple sources are many, we will not list these but direct the reader to appropriate references [Luo and Kay, 1989, Waltz and Buede 1986,]. However, it is worth mentioning two very fundamental advantages viz. *redundancy* and *complementarity* [Luo and Kay, 1989] as these form a part of the guiding principles in what follows.

When a number of sensors detect the same feature of the target, possibly with different fidelity, we say that the collected information has redundancy. Integration of redundant information reduces overall uncertainty. Furthermore, a system capable of accumulating redundant information will not be drastically paralysed owing to accidental failures of any particular sensor.

Collecting complementary information, on the other hand, implies collecting qualitatively different information regarding the enemy. Integration of such information provides a holistic appreciation of the environment. Although redundancy is desirable when the goal is the reliable estimation of a certain parameter, complementarity is useful in performing global assessments.

Bayesian networks, since their inception, have shown great promise in performing multisensor data fusion [Waltz and Llinas 1990]. We will illustrate both the above types of integration in the illustrative network that we construct for situation assessment. We assume that there are external modules which receive sensor data and make the data available as input evidence to the network.

2.5 A Scenario for Naval Anti Surface Warfare

To endow the analysis with a degree of concreteness, we construct a scenario of naval anti-surface warfare. This scenario is based upon *operation Dardanelles;* an approved naval exercise [ANZAC scenario 1997]. The situation assessment is then undertaken with respect to this scenario.

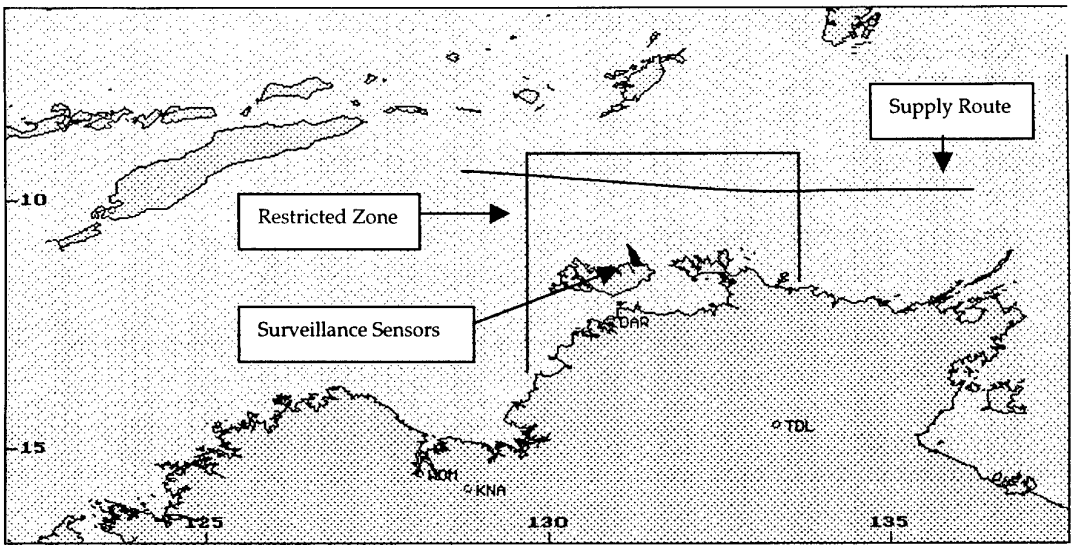

Figure 2.1. Map showing the *restricted zone, supply route,* and the position of *surveillance facilities.*

The scenario (Figure 2.1) develops during a period of developing hostility between Blue and Orange forces. From intelligence gathered during early stages of conflict it is apparent that the enemy military commanders view the communications and surveillance facilities of the Blue forces, based somewhere on the northern coast, as a significant target. These facilities are able to observe supply routes through which the Orange forces get fuel supply from a neighbouring country with which it has recently signed a secret defence pact disguised as a trade pact. This third country is to be treated as neutral and Blue forces will not open fire on its tankers. As the situation worsens the Blue forces designate an area covering land and sea bounded by some grid points XXX, and containing the communications facilities, as a *restricted zone.* The friendly government has indicated that any military activity or any supply transported through the designated area that fosters military activity will be treated as hostile activity. The fuel supply is vital to all enemy future plans, the Orange forces would therefore like to defend the supply route that pass through the restricted zone. It would also like to destroy the surveillance facilities but that would be attempted only when hostilities break out openly.

HMAS ANZAC is assigned to take command of a Blue task force in detecting, tracking and attacking if necessary all enemy surface units in the restricted zone. The Officer in Tactical Command (OTC) is the Commanding Officer, HMAS ANZAC.

The following constitute the Blue task force

- HMAS ANZAC
- One Patrol Boat FCBP
- One Maritime Patrol Aircraft (MPA)
- One Helicopter
- One F111 (Maritime strike aircraft)

Initial intelligence suggests that the orange forces would use one or more of the following surface crafts for intrusion into the restricted zone.

- A major fleet unit; a MOD KASHIN FFG
- One Patrol Boat
- A communication ship

In addition, neutral supply vessels are expected to attempt passing through the restricted zone.

Rules of engagement for HMAS ANZAC will be along the following lines. Initiate action to detect any Orange surface unit or neutral tankers, inside the restricted zone using MPA, Helicopter and own sensors. Also use third party information received from offshore surveillance facilities. Once detected conduct continuous tracking using MPA, Helicopter, etc. ensuring that the tracking units are outside enemy fire range. Provide enemy unit(s) coordinates to other units in the Blue task force. In particular, provide targeting information to Blue Maritime Strike Aircraft (MSA) F111. If a tanker belonging to the neutral country is detected in the restricted zone, intercept it and send it back to the port of origin. If Orange surface units are detected then inform Joint Operation Room (JOR) of the situation. Attack only when directed by the MHQAUST from the JOR. When an attack becomes necessary coordinate attack with other Blue units. CO HMAS ANZAC is to pursue attack until the expected *probability of kill* is achieved.

Intelligence suggests that the Orange forces course of action will be along the following lines. At the outset Orange forces would prefer to get the supply ships through the restricted zone undetected. If the ships get detected the Orange forces will wait until the Blue forces take any action. They have correctly inferred that Blue forces will be reluctant to attack a neutral country's tankers. In case there is a build up of Blue forces or the supply tankers get sent back, the Orange forces will actively defend the supply route. If the hostilities escalate or if the Orange force commanders perceive that the Blue surveillance facilities pose a significant threat to overall Orange war activity they will mount an attack on the Blue task force with the ultimate aim of destroying the facilities.

2.6 An Illustrative Bayesian Network

We now construct a Bayesian network (Figure 2.2) to assess a variety of situations as are likely to be generated by the scenario constructed above. First of all we decide upon a set of hypotheses that adequately represents possible *enemy intentions*. Assume that intelligence gathered so far and an analysis of the pattern of Orange forces past behaviour provides us with the following set options:

- *Passive*: Maintain remote surveillance of the restricted zone with sensor platforms that are well out of range of Blue fire power. Ask the neutral country to commence fuel supply and assume that Blue forces will not interfere with the activities of an apparently neutral country.
- *Defensive*: Conduct active reconnaissance and maintain a defensive presence to guard the supply routes against Blue forces interference.
- *Offensive*: Mount naval attack on Blue surface units with the intention of neutralising them and destroying Blue offshore surveillance facilities.
- *Not Modelled*: Other possible enemy options not covered above.

Enemy intention directly influences *enemy activity,* this can take one of the following forms:

Enemy activities:

- Logistics – procuring fuel supply from the neutral country.
- Reconnaissance of the restricted zone.
- Conducting EW
- Securing supply route
- Mounting naval attack
- Inactive when severely damaged

The enemy surface units likely to be deployed are detailed in the scenario. We add the supply tankers from the neutral country to that list.

Identification of the enemy unit and gathering information about its activity is achieved by surveillance through a number of sensors and reconnaissance. These information gathering devices observe the following information variables:

Enemy Vessel Type:

- FFG
- FCPB
- Communication vessel
- Oil Tanker

Position of the enemy unit:
- Outside the restricted zone
- Near the boundary of the zone
- Near supply route
- Well forward

Mobility of the enemy unit:
- Immobile
- Slow parallel (parallel to the northern boundary of the restricted zone)
- Slow forward
- Slow backward
- Rapid parallel
- Rapid forward
- Rapid backward

Communication activities of enemy unit:
- Maintaining radio silence
- Communicating with base
- Jamming (EW)

Evidence to the network is supplied through the sensor and reconnaissance nodes. For example, the evidence regarding the *position of enemy unit* is supplied through the nodes *Sensor Position Int* and *Recon Position Int*. Here we have assumed that there is only one sensor unit and one reconnaissance unit to detect enemy position; a greater number of units would be represented by additional nodes of similar nature. The links from the *position* node to the corresponding *sensor* and *reconnaissance* nodes capture the confidence that the analyst places on the reliability of these detection units. If the sensor unit detects the enemy *near the supply route*, the *sensor* node is instantiated to this state. The network utilises this evidence to update the probability distribution over the states of the *position* node. This updated distribution will not necessarily assign a value *100* to the state *near the supply route* in the *position* node.

As shown in Figure 2.2, the parameters *vessel type, position* and *mobility* are each detected through sensors and reconnaissance. The collected information therefore has redundancy. The network integrates the redundant information for each of the above parameters to reduce the overall uncertainty in information gathering. If the nodes *Sensor Position Int* and *Recon Position Int* provide conflicting information, the network utilises the conditional probabilities attached to the links from the *position* node to these nodes to integrate the information and work out a probability distribution over the states of the *position* node.

The network also integrates complimentary information. For example the state of *activity* is determined by integrating information regarding *vessel type, communication activity, position* and *mobility*.

Depending upon the state of enemy intention and its activity the network provides a probability distribution over possible courses of action - *Blue Courses of Action:*

- Dispatch patrol boat to intercept neutral vessel
- Dispatch F111 for small scale attack
- Mount full scale attack with HMAS ANZAC and F111

This is in addition to surveillance, reconnaissance and communication with JOR and other Blue units that would continue to be carried out routinely.

2.6.1 A Bayesian Network for Situation Assessment

The following Bayesian network was constructed using NETICA [Netica 1998]. In Appendix C we demonstrate the use of this Bayesian network in belief updating for a developing situation.

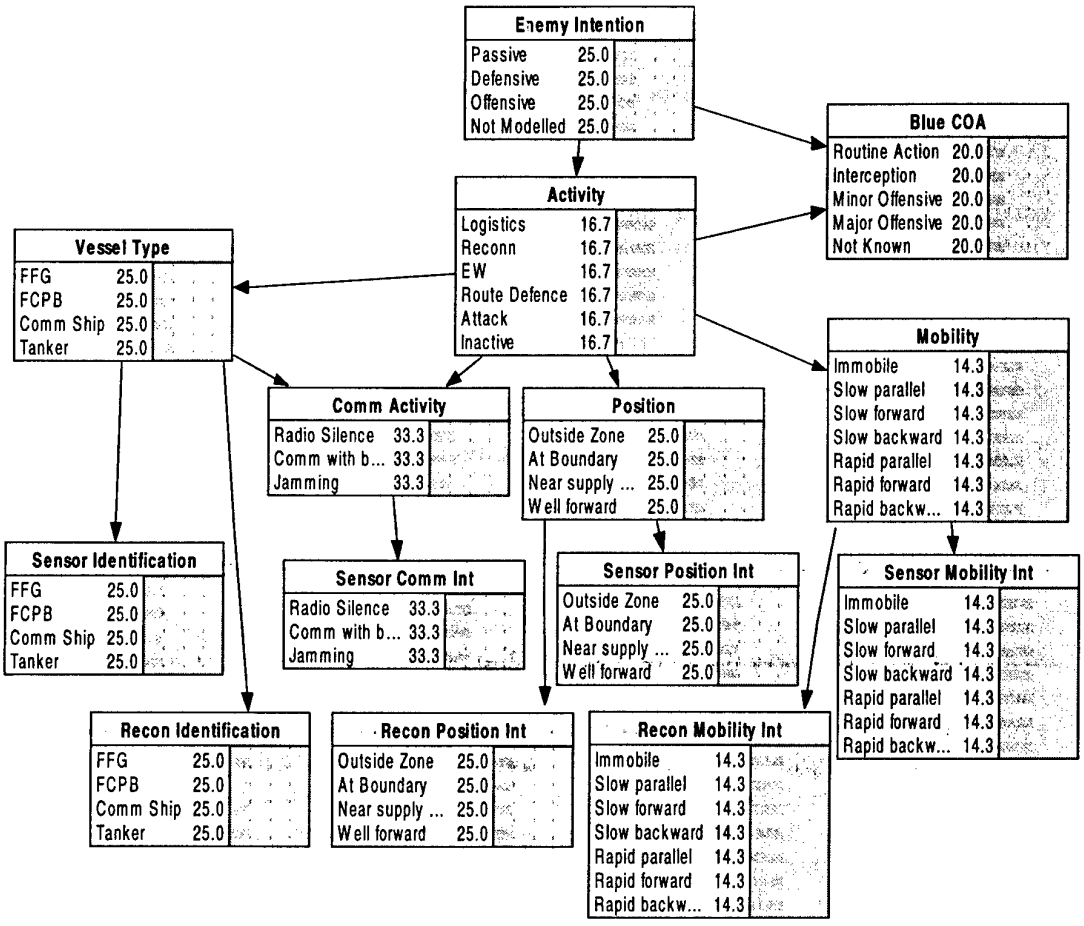

Figure 2.2. A Bayesian network for situation assessment.

3. Measures of Uncertainty and its Flow in a Bayesian Network

As the preceding section illustrates, a Bayesian network is primarily used to update the probability distribution over the states of a hypothesis variable; a variable which is not directly observable. This probability distribution then helps a decision maker in deciding upon an appropriate course of action. The question now arises as to how *effective* is such a process. In particular it would be useful to be able to answer the following:

- Every act of information gathering consumes some amount of resources. One can therefore ask if a particular stage of hypothesis updating was worth the amount of resources spent in information gathering for that stage.
- With respect to a particular Bayesian network how effective is a particular information gathering device?
- Further, a Bayesian network is nothing but a tool in aid of decision making. Given a situation and a corresponding Bayesian network, how effective is this network as a decision aid tool; does it remain effective as the situation changes with time?

A complete answer to the above questions and the related ones that will crop up once we start probing, is beyond the scope of this present work. What we propose to do is establish a methodology that can be used to find the answers. To show that the methodology actually works we will then find some initial answers and compute a few measures that will be sufficient for the present.

3.1 A Measure of Uncertainty

To answer questions regarding effectiveness, the central theme to bear in mind is that the uncertainties that reside in various variables taken individually or taken in groups form key factors influencing any decision making exercise. The first thing to do would therefore be to fix a measure of uncertainty associated with any random variable.

Let X be a discrete random variable taking the values $\{x_1,\ldots, x_n\}$ with probabilities $P\{X=x_i\} = P(x_i)$. We have all along stressed the point of view that our interpretation of the probability here is the subjective or the Bayesian interpretation. This holds that a probability measure is a measure of our belief based on our current state of knowledge. The variable X will assume one of the n alternative values $\{x_1 \ldots, x_n\}$, if the corresponding experiment were performed. We do not know exactly which alternative will materialise, our knowledge only provides us with a measure that x_i will occur with the likelihood $P(x_i)$. In a sense, this probability distribution is a measure of our *current knowledge*, the knowledge that was used to assign the probabilities in the first place. However this knowledge is not enough to pin down that particular alternative which

will materialise when the experiment is performed. This corresponds to our *current uncertainty*.

> *A measure of the amount of current uncertainty is a measure of the amount of information that would be required in addition to the current knowledge to specify which particular alternative will occur.*

Equivalently, a measure of our current uncertainty is a measure of the amount of information that will be *acquired* when one performs the experiment and determines the particular alternative. Hence a measure of uncertainty would be determined by the number of available alternatives and the probability distribution $P(\cdot)$. If we further stipulate that the measure of uncertainty be given by the logarithmic measure of the number of alternatives then it can be shown that (Caves and Fuchs 1996) the average uncertainty associated with the random variable X is given by

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) \tag{3.1}$$

This is just the *Shannon entropy* of the random variable X. If we use logarithms to base 2, as we shall in all our considerations henceforth, the unit of entropy will be a bit. It measures the *average* information required in addition to the current knowledge to specify a particular alternative. If our current state of knowledge is complete, i.e. if we know that X will assume the value x_1, say, then $P(x_1) = 1$ and it follows that $H(X) = 0$. If our current state of knowledge is *total ignorance* then we will not be able to distinguish between various alternatives, this leads to the uniform probability distribution $P(x_1) = 1/n$. In this case, the additional information required to pin down an alternative will be maximum. This is precisely what happens when we evaluate $H(X)$ with uniform distribution, it takes the value $\log n$ the maximum possible value. In general therefore, $H(X)$ provides a measure of the amount of information required to remove the ignorance expressed by the probability distribution $P(x_1)$.

In the analysis that follows it would sometimes be useful to adopt another equivalent interpretation. We can view $H(X)$ as a measure of the spread of the probability distribution $P(x_i)$. In this respect the entropy plays a role similar to the concept of variance in statistics. However, in many respects $H(X)$ is a superior measure of the spread. Variance can only be defined if the values $\{x_1, \ldots, x_n\}$ assumed by X are real numbers. This is far too restrictive for our purpose, where X may stand for *enemy intention* taking values from the set { *offensive, defensive, logistics etc.*}. $H(X)$ as a measure of the spread is unencumbered by the intrinsic nature of X, its only dependence is on the probability distribution and therefore reflects the spread more faithfully.

Normally, one studies the propagation of evidence as it ascends from the information variables to the hypothesis variables, higher up in the causal link, in any Bayesian network. Essentially therefore, the network is viewed as an *information channel*. We want to obtain a better understanding of this flow of information, and the strategy adopted for this is very simple – apply the concepts and results from information theory to Bayesian networks. The necessary material from information theory can be

found in any standard text [eg., Cover and Thomas 1991 - will henceforward be referred to as **CT**], and we start with a few definitions.

Joint Entropy

Let X and Y be two random variables with a joint distribution given by $P(x,y)$. The joint entropy $H(X,Y)$ is given by

$$H(X,Y) = -\sum_i \sum_j P(x_i, y_j) \log P(x_i, y_j) \tag{3.2}$$

Conditional Entropy

Let X and Y be two random variables. The conditional entropy $H(X \mid y_j)$ of X given that Y has assumed the value y_j is given by

$$H(X \mid y_j) = -\sum_i P(x_i \mid y_j) \log P(x_i \mid y_j) \tag{3.3}$$

The conditional entropy $H(X \mid Y)$ of X with respect to Y is the expected value of the measure of uncertainty in X when it is known that Y has a particular value. Hence clearly it has the following form.

$$H(X \mid Y) = \sum_j P(y_j) H(X \mid y_j) = -\sum_j P(y_j) \sum_i P(x_i \mid y_j) \log P(x_i \mid y_j) \tag{3.4}$$

$H(X \mid Y)$ quantifies, on the average, the remaining uncertainty regarding X when it is known that Y has assumed a certain value.

3.1.1 Information Equivalence of Bayes Theorem

Since evidence propagates along the network according to the dictates of Bayes theorem, let us first derive a result that has the same meaning for entropy as Bayes theorem has for probability. The joint probability $H(X,Y)$ for two random variables X and Y can be shown to satisfy the relationship [CT]

$$H(X,Y) = H(X \mid Y) + H(Y). \tag{3.5}$$

From the commutative property of logic it follows that the conjunction X *and* Y is equivalent to the conjunction Y *and* X. Hence $H(X,Y) = H(Y,X)$, thus we conclude:

$$H(X \mid Y) = H(Y \mid X) + H(X) - H(Y) \tag{3.6}$$

Following Bayesian tradition, we say that $H(X)$ represents the initial or *prior* uncertainty in X, i.e. the uncertainty in our knowledge about X when no other information is available. While $H(X \mid Y)$ represents the updated or *posterior* uncertainty, i.e. the average remaining uncertainty in X when we have evidence in hand that Y has assumed a certain value. Equation (3.6) therefore shows the relation between the *prior* entropy and the *posterior* entropy. Further it can be proved that [CT]

$$H(X) \geq H(X \mid Y), \tag{3.7}$$

with equality if and only if X and Y are independent. Hence belief updating decreases entropy, on the average, a conclusion in keeping with the trends of Bayesian analysis. This also reassures us that our choice of measure is an appropriate one.

3.2 Effect of Evidence on Hypothesis in a Bayesian network

An important point that must be kept in mind is that the reduction in entropy as mentioned in (3.7) is true only on the average. Like all averages this glosses over a very significant fact which is the following: if the random variable Y is known to have assumed a particular value y then the conditional entropy $H(X \mid Y=y)$ will not always be less than $H(X)$, it may at times be greater than $H(X)$. In other words, arrival of some particular new information can sometimes increase the uncertainty (in situation assessment for example) - although, on the average information always reduces uncertainty.

Consider a Bayesian network and let X with values $\{x_1,\ldots x_n\}$ be a hypothesis variable while $E = \{e_1,\ldots e_d\}$ be an information variable. If now information arrives that E has been found to be in the state corresponding to e_1, then this propagates through the network and changes the probability distribution of X to $P(x_i \mid e_1)$. The entropy of the hypothesis variable now becomes $H(X \mid e_1)$. Now whether this posterior entropy is greater than, or equal to, or less than the prior entropy $H(X)$ depends upon the following two factors:

1. The prior probabilities $P(x_i)$
2. The knowledge stored in the links of the network in terms of the conditional probabilities. These conditional probabilities govern the reasoning system that the network utilises to derive inferences.

We are interested in finding out how effective the network's reasoning system is in employing evidence to reduce uncertainty. To do this, the effects of the prior probabilities must first be decoupled from the effects of the reasoning system. We adopt the following simple but very effective strategy for this. We assume that we have no prior knowledge of the hypothesis variable, and then examine the capacity of the evidence to reduce this total ignorance or maximum uncertainty.

If nothing is known about the hypothesis X with values $\{x_1,\ldots,x_n\}$, then according to the *principle of maximum entropy* [Jaynes 1988] we must have $P(x_i) = 1/n$ for all i and the prior entropy would be $H(X) = \log n$. When the evidence that E has been found to be in the state e_1 arrives, this entropy gets reduced by an amount

$$\log n - H(X \mid e_1) \tag{3.8}$$

In (3.8) the posterior probabilities $P(x_i \mid e_1)$ are calculated using the prior probabilities $P(x_i) = 1/n$ for all i. Furthermore, since $\log n$ is the maximum entropy possible for X the expression in (3.8) is always positive. We therefore call the expression in (3.8) *the uncertainty reducing capacity of the evidence 'E is in state e_1'*. If $\{\varepsilon_1, \varepsilon_2,\ldots,\varepsilon_s\}$ are a set of evidence obtained through various sources then we can use the above formalism to calculate the uncertainty reducing capacity of the set as $\log n - H(X \mid \varepsilon_1, \varepsilon_2,\ldots,\varepsilon_s)$. Moreover this expression can be used to compare the uncertainty reducing capacity of different sets of evidence.

The question now arises as to how would one calculate the uncertainty reducing capacity of some evidence using a Bayesian network. The following three cases arise:

(a) First suppose that the hypothesis variable X is a root node. We assign a probability distribution to X such that all its possible values are equally probable; we do the same to all other root nodes. After the network has reached an equilibrium with these initial conditions we instantiate one or more information variables to their corresponding evidence values. The posterior probabilities $P(x_i \mid \varepsilon_1, \varepsilon_2,\ldots,\varepsilon_s)$ get calculated when the network reaches equilibrium again and we use these to calculate the uncertainty reducing capacity of the evidence.

(b) If the hypothesis variable X is not a root node then we delete from the network the parents of X and carry out the process described in (a).

(c) It may also happen that an evidence variable influences X through one of its parents Y say. In such cases deleting the parents will render the evidence and hypothesis disconnected. Since the evidence has direct influence on Y and only indirect influence on X, we treat Y as the primary hypothesis and fix its prior probabilities instead. In other words, if Y is a root node then we make all its possible states equally likely as in (a) and if Y is not a root node we make it one in the manner discussed in (b). Since Y is a parent of X fixing prior probabilities of Y would automatically fix the prior probabilities of X.

Appendix D.1 illustrates these calculations with an example.

Finally the average effect of any evidence variable E on a hypothesis variable X can be calculated by instantiating E to each of its possible values successively and propagating the effects to X. This will allow us to calculate $H(X \mid E)$ using (3.4) leading to the average uncertainty reducing capacity of E which is $(\log n - H(X \mid E))$.

3.3 Effect of belief updating

Although reduction of uncertainty is the primary motivation for gathering information, there are other subtle features of updated information that play a crucial role while deciding upon a subsequent course of action. Consider again the hypothesis variable X which models various states of some situation. If the probability distribution is such that $P(x_i)$ is very near to 1 for some x_i then we know with reasonable confidence that the state of the situation is given by x_i. This is however rarely the case. When the probability distribution is spread out over many states of the hypothesis, it is the nature of the spread, or the form of the probability distribution function P that is very important. The form of P *dictates our belief* regarding the peculiarities of the situation being observed. In a dynamic environment the nature of the situation would change with time. Our observations may never give us an exact description of the situation but the sequence of probability distributions we obtain, as we go on updating information, should inform us how the situation is changing.

To formulate the above considerations quantitatively, consider the following development. Suppose at time t_1 the probability distribution over the hypothesis X is P_1. A batch of sensor observations arrives at a subsequent time t_2 leading to a new distribution P_2; let us assume that the updated information faithfully captures the change in the situation. We now ask - what would be the inefficiency in the decision making process if the information had not been updated? With no update all decisions at time t_2 would be based on the *perceived* distribution P_1. Since the actual distribution is P_2 the uncertainty in our perception would be given by, $-\sum_i P_2(x_i)\log P_1(x_i)$. This is different from $-\sum_i P_2(x_i)\log P_2(x_i)$ which would measure the uncertainty if our perceptions were based on the updated information. Hence the increase in uncertainty due to outdated perception is given by

$$-\sum_i P_2(x_i)\log P_1(x_i) + \sum_i P_2(x_i)\log P_2(x_i) \tag{3.9}$$

What we have obtained is the *Kullback distance* between the distributions P_2 and P_1 which is generally denoted by $D(P_2\|P_1)$. It can be shown that $D(P_2\|P_1)$ is always positive [CT], and this is very interesting - for the amount of uncertainty associated with the updated distribution P_2 may actually be greater than the amount of uncertainty associated with the previous distribution P_1. In other words obtaining new evidence may not necessarily decrease uncertainty, as we have said before, but not updating information is always inefficient. $D(P_2\|P_1)$ therefore correctly measures the worth of the updated information. The amount of resources spent in reconnaissance and surveillance must therefore be weighed against the value of $D(P_2\|P_1)$ to determine their effectiveness.

In Appendix D.2 we give an example and calculate $D(P_2\|P_1)$ for a stage of situation assessment using the Bayesian network to update probabilities.

3.4 Sensor Effectiveness

To analyse how the reliability of information gathering devices affect the process of inference making in a Bayesian network, we first define the concept of mutual information [CT].

Mutual Information

Given two random variables X and Y, the mutual information $I(X;Y)$ is defined as

$$I(X;Y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \tag{3.10}$$

It can be shown that

$$I(X;Y) = I(Y;X) \text{ and} \tag{3.11}$$

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{3.12}$$

The mutual information $I(X;Y)$ is a measure of the amount of information that Y contains about X. In other words $I(X;Y)$ provides the uncertainty reducing capacity, on the average, of the random variable Y with respect to the uncertainties in X.

Let us now consider the following simple formulation which captures all the essential factors required for an analysis of sensor effectiveness.

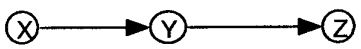

Figure 3.1. A Bayesian network; X denotes hypothesis variable, Y denotes information variable and Z represents a sensor for obtaining information for Y.

Here X is the hypothesis variable assuming values $\{x_1,\ldots x_n\}$. This hypothesis is evaluated through information provided by the information variable Y. The variable Z represents a sensor which feeds information to the information variable. Y and Z have similar states as Z gathers information required by Y – for example in the case of the situation assessment Bayesian network depicted in Figure 2.2, Y can represent the *position* while Z represents *sensor position int*. Let us represent the values assumed by these variables as follows:

$$Y : \{y_1,\ldots,y_d\}$$

$$Z : \{z_1,\ldots,z_d\} \tag{3.13}$$

If Z were a perfect sensor then instantiating Z to z_l would imply that the information variable Y is in the corresponding state y_l with probability *1*. In general we have

$$P(z_i \mid y_j) = \delta_{ij} \quad i,j = 1 \ldots d \tag{3.14}$$

The marginal probabilities $P(y)$ and $P(z)$ are therefore related as below

$$P(z_i) = \sum_{j=1}^{d} \delta_{ij} P(y_j) \tag{3.15}$$

Hence for reliable sensors we must have very small values for δ_{ij} if $i \neq j$, and δ_{ii} very near to *1*, where $\delta_{ii} = 1 - \sum_{\substack{i=1 \\ i \neq j}}^{d} \delta_{ij}$.

For a perfect sensor we have

$$\begin{aligned} \delta_{ij} &= 0 \qquad \text{if } i \neq j \\ &= 1 \qquad \text{if } i = j \end{aligned} \tag{3.16}$$

In other words we consider the link between Y and Z as a *noisy* information channel. The noise factors are represented by the conditional probabilities $P(z_i \mid y_j)$. The other link between the variables X and Y quantified by the conditional probabilities $P(y_i \mid x_j)$ represents the knowledge for inference making about the hypothesis.

If X has prior probabilities $P(x_i)$, the initial uncertainty in the hypothesis is given by $H(X) = -\sum_i P(x_i) \log P(x_i)$. To calculate the uncertainty reducing potential of the sensor, let us consider the case when the sensor indicates that Z has been found to be z_s. The posterior uncertainty in the hypothesis now becomes

$$H(X \mid z_s) = -\sum_i P(x_i \mid z_s) \log P(x_i \mid z_s) \tag{3.17}$$

Here

$$\begin{aligned} P(x_i \mid z_s) &= \frac{\sum_k P(x_i, y_k, z_s)}{P(z_s)} = \frac{\sum_k P(z_s \mid y_k) P(y_k \mid x_i) P(x_i)}{P(z_s)} \\ &= \frac{\sum_k \delta_{sk} P(y_k \mid x_i) P(x_i)}{\sum_k \delta_{sk} P(y_k)} \end{aligned} \tag{3.18}$$

For a perfect sensor using the properties (3.16) we get

$$P(x_i \mid z_s) = \frac{P(y_s \mid x_i) P(x_i)}{P(y_s)} = P(x_i \mid y_s) \tag{3.19}$$

Hence for perfect sensors we have

$$\begin{aligned} & H(X \mid Z) \\ & \quad = -\sum_s P(z_s) \sum_i P(x_i \mid z_s) \log P(x_i \mid z_s) \\ & \quad = -\sum_s P(y_s) \sum_i P(x_i \mid y_s) \log P(x_i \mid y_s) \\ & \quad = H(X \mid Y) \end{aligned} \tag{3.20}$$

Now the average uncertainty reducing capacity of the sensor Z is given by $I(X;Z)$. Using the independence property of the Bayesian network i.e. the property that X and Z are independent given Y it can be easily shown that

$$I(X;Z) \leq I(X;Y). \tag{3.21}$$

However for a perfect sensor it follows from (3.20) that $I(X;Z) = I(X;Y)$. In other words a perfect sensor will on the average reduce the uncertainties in X by an amount $I(X;Y)$, as the reliability of the sensor decreases its uncertainty reducing capacity also decreases and the amount by which it decreases is given by

$$I(X;Y) - I(X;Z). \tag{3.22}$$

We shall use the numerical value of expression in (3.22) to measure the efficiency of the sensor with regard to its uncertainty reducing capacity – *the less the value is, the more effective is the sensor.*

The above analysis is easily generalised to any Bayesian network. Numerical calculations will proceed along the following two steps:

i. Fix the prior probabilities $P(x_i)$, by making X a root node as discussed in section 3.2.

ii. To calculate $I(X;Z)$ start by instantiating Z to its various possible states. The resulting probability distribution of X obtained for each instantiation would allow calculation of the conditional entropy $H(X \mid Z)$. Formula (3.10) then gives $I(X;Z)$. A similar procedure will compute $I(X;Y)$.

Appendix D.3 calculates the effectiveness of the position detecting sensor in relation to the hypothesis *Enemy Intentions* in the context of situation assessment.

3.5 Effectiveness of a Bayesian Network

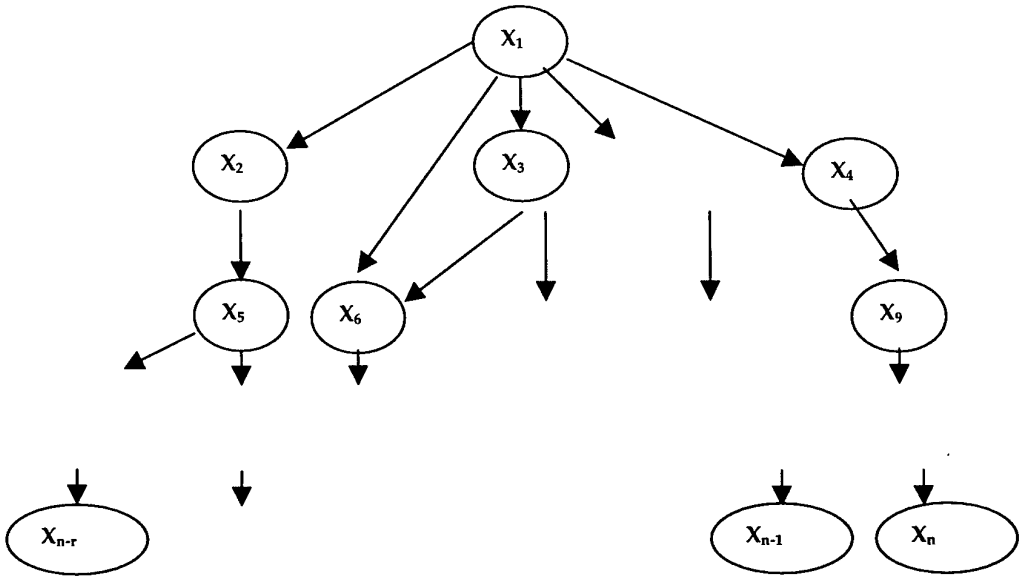

Figure 3.1. A general Bayesian network with one hypothesis variable and a number of information variables.

Consider a general Bayesian network as shown above having n nodes. The root node X_1 represents the hypothesis variable while the nodes with no children represent the $r+1$ information variables X_{n-r}, ..., X_{n-1}, X_n. The rest are intermediate nodes which help propagate evidence from the information variables to the hypothesis variable. Once the n nodes are fixed and all the links between the nodes are fixed, the network is structurally fixed. If we further specify all the conditional probabilities that quantify the links, then the network is *functionally specified*. The network can now be used to update the probability distribution function over the hypothesis variable when new evidence are injected through the information variable.

A note on notations: suppose the random variable X_k assumes values in the set $\{x_{k_1},...,x_{k_m}\}$ then for the sake of clarity in notation we will use $-\sum_{x_k} P(x_k)\log P(x_k)$ to denote $H(X_k)$ where it is understood that the summation is over the values in the set $\{x_{k_1},...,x_{k_m}\}$.

The *effectiveness of the network* is determined by its ability to utilise information to update belief in the hypothesis. How well it performs this function, of course, depends upon the functional specification which specifies the degree of influence that the information variables have over the hypothesis variable. Hence a measure of

effectiveness can be obtained by obtaining a measure of this influence and the latter is measured by the mutual information function:

$$I(X_1;X_n, \ldots, X_{n-r}) = H(X_n, \ldots, X_{n-r}) - H(X_n, \ldots, X_{n-r} \mid X_1) \tag{3.23}$$

To understand the nature of the mutual information we recall the fact that given two random variables X and Y the mutual information $I(X;Y)$ is a concave function of $P(x_i)$ for fixed $P(x_i \mid y_j)$ [CT]; we here obtain a similar result for Bayesian networks of the type shown in figure 3.1. Let us examine the first term in the right hand side of (3.23):

$$H(X_n,\ldots,X_{n-r}) = -\sum_{x_n,\ldots,x_{n-r}} P(x_n,\ldots,x_{n-r}) \log P(x_n,\ldots,x_{n-r}) \tag{3.24}$$

Where the probabilities in (3.24) can be written as

$$\begin{aligned} P(x_n,\ldots,x_{n-r}) &= \sum_{x_1,\ldots,x_{n-r-1}} P(x_n,\ldots,x_2,x_1) \\ &= \sum_{x_1,\ldots,x_{n-r-1}} P(x_n \mid pa(x_n)) \times P(x_{n-1} \mid pa(x_{n-1})) \times \cdots \times P(x_2 \mid pa(x_2)) \times P(x_1) \end{aligned} \tag{3.25}$$

Here $pa(X_i)$ denote the parent set of the node X_i and we have used the conditional independencies inherent in the structure of the Bayesian network. As the conditional probabilities $P(x_i \mid pa(x_i))$ are fixed it follows that $P(x_n,\ldots,x_{n-r})$ is a linear function of $P(x_1)$. Hence, $H(X_n,\ldots,X_{n-r})$ which is a concave function of $P(x_n,\ldots,x_{n-r})$ is a concave function of $P(x_1)$.

Further, the second term in the right hand side of (3.23) is

$$H(X_n,\ldots,X_{n-r} \mid X_1) = -\sum_{x_1} P(x_1) \sum_{x_n,\ldots,x_{n-r}} P(x_n,\ldots,x_{n-r} \mid x_1) \log P(x_n,\ldots,x_{n-r} \mid x_1) \tag{3.26}$$

We have

$$P(x_n,\ldots,x_{n-r} \mid x_1) = \frac{P(x_n,\ldots,x_{n-r},x_1)}{P(x_1)}$$
$$= \frac{\sum_{x_2,\ldots,x_{n-r-1}} P(x_n,\ldots,x_2,x_1)}{P(x_1)}$$
$$= \frac{\sum_{x_2,\ldots,x_{n-r-1}} P(x_n \mid pa(x_n)) \times P(x_{n-1} \mid pa(x_{n-1})) \times \cdots \times P(x_2 \mid pa(x_2)) \times P(x_1)}{P(x_1)}$$
$$= \sum_{x_2,\ldots,x_{n-r-1}} P(x_n \mid pa(x_n)) \times P(x_{n-1} \mid pa(x_{n-1})) \times \cdots \times P(x_2 \mid pa(x_2)) \quad (3.27)$$

Hence $P(x_n,\ldots,x_{n-r} \mid x_1)$ is fixed once all the conditional probabilities $P(x_i \mid pa(x_i))$ are fixed it follows from (3.26) that $H(X_n,\ldots,X_{n-r} \mid X_1)$ is a linear function of $P(x_1)$.

$I(X_1;X_n, \ldots, X_{n-r})$ is the difference between a concave function of $P(x_1)$ and a linear function of $P(x_1)$ and is therefore a concave function of $P(x_1)$.

If the hypothesis variable admits h alternatives ie. $X_1 = \{x_{1_1},\ldots,x_{1_h}\}$ then the probability distribution function $P(x_1)$ is given by the h numbers $\{P(x_{1_1}),\ldots,P(x_{1_h})\}$. The set of all such probability distribution functions determine a h-1 dimensional simplex in R^h defined by the relations

i. $0 \le P(x_{1_i}) \le 1 \qquad i = 1,2,\ldots,h$ (3.28)

ii. $\sum_{i=1}^{h} P(x_{1_i}) = 1$ (3.29)

$I(X_1;X_n, \ldots, X_{n-r})$ is a concave function defined over this simplex. Hence there exists a connected subset S of the simplex over which the mutual information is constant and this constant value is the global maximum; if S is a single point then $I(X_1;X_n, \ldots, X_{n-r})$ attains the unique global maximum at this point.

The prior probability distribution $P(x_1)$ over the hypothesis variable, at any stage of decision making, reflects the nature of the situation at that stage. As the situation changes the probability distribution varies over the simplex discussed above. What the conclusion reached above suggests is that if the situation is such that $P(x_1)$ belongs to the set S then the functionally specified network makes maximum utilisation of the information gathered. As $P(x_1)$ wonders out and recedes away from S reflecting a manner of evolution of the situation the ability of the network to exploit new evidence decreases. To bring back optimal performance one has to change the functional specifications so that with respect to the new network $P(x_1)$ is again a point within the corresponding set S or a point near the boundary of S. This change can involve either or both of the following:

i. Change the set of observables so that the information gathered has a greater degree of relevance to the prevailing situation.

ii. Change the intermediate nodes and therefore the links in the network so that the channel of evidence propagation is more relevant to the prevailing situation

The concavity of $I(X_1;X_n, \ldots, X_{n-r})$ is a very significant fact. It suggests that the network is most effective with respect to a unique situation or a set of situations, corresponding respectively to a unique point or a connected region in the probability simplex. Recall that we have always laid stress on the subjective nature of the network. The network reflects the personalistic reasoning process of a decision maker. We now answer the following crucial questions. Is this subjective process *efficient* and *internally consistent*? Given a hypothesis state, i.e. a situation, has the decision maker efficiently reasoned through the causal chain to identify the appropriate observables? He or she has, if the mutual information $I(X_1;X_n, \ldots, X_{n-r})$ attains its maximum at a point in the probability simplex close to the point representing the given situation. If on the other hand, these two points are far apart then the *subjective process* of obtaining the network has not been efficient and is possibly internally inconsistent.

4. Conclusion

The main driving force behind this work has been a desire to find suitable tools to model uncertainties in command and control. The graphical techniques of Bayesian networks, as we have demonstrated, provide a rich tool to comprehend and analyse uncertainties. The pictorial display of the model as a graph facilitates easy understanding and therefore would be of great help in rapid model development.

The framework of Bayesian networks divides the model development process into two parts decoupling the qualitative aspects from the quantitative ones. This enables the user to first concentrate on building the causal structure of the network without worrying about the probabilistic aspects. All that is required is a clear understanding of the causes and their effects regarding the C3I problem under consideration, and this is what the commanders are good at. The next step is quantitative – assigning conditional probabilities to the links. In the section dealing with situation assessment we have advocated that the commander's perception or subjective knowledge of the situation be exploited to obtain the conditional probabilities, and as pointed out there, this is a problem requiring further research.

Other aspects which require further research as discussed earlier are:

- Given a real world scenario how to generate a set of hypotheses that capture possible enemy intentions and
- How to obtain the prior probabilities to initiate the network?

The information theoretical techniques developed in section 3 to analyse the flow of uncertainty in a Bayesian network opens up a new line of research. We have demonstrated how these techniques can be used to answer various questions regarding effectiveness. However the trend of the analysis, in particular the concavity of the mutual information as proved in section 3.5, indicates that these techniques can also be fruitfully used to investigate problems related to learning.

In a modelling and simulation environment, Bayesian networks need to be integrated with other simulation tools. Essentially this integration, as with all such integrations, should make it possible for the network to interact with other components of simulation i.e. access relevant data as input and produce probability distributions in a manner that can be accessed by other components. The input to a network will be in the following two forms:

i. prior probability distributions as input to all hypothesis variables or root nodes and

ii. data in the form of evidence as input to information variables.

The output(s) are all the probability distributions over all the nodes.

For example, the situation assessment network discussed before would require data regarding enemy position and mobility etc. from modules that simulate the physical environment. All the probability distributions produced by the network can be used by other modules which deal with one or other aspects of determining an appropriate *course of action*.

5. Acknowledgments

The author wishes to thank Mark Nelson for being very helpful in providing a number of references – to Joanne Nicholson and Rod Staker for sharing their knowledge regarding Bayesian networks. Thanks are also due to Lin Zhang who provided the documents based upon which the scenario is constructed and to Noel Haydon whose maps helped illustrate the scenario. Finally thanks are due to Mike Davies who suggested the Bayesian network approach for modelling and simulation.

6. References

Almond, R. G., (1996), Software for Manipulating Belief Networks, http://bayes.stat.washington.edu/almond/belief.html

ANZAC Scenario (1997), ANZAC Ship Acceptance Test Plan Volume 3 Section 6 – Run Plans and Geometries. *ADF Publications.*

Berger, J. O., Boukai, B., and Wang, Y., (1997) Unified Frequentist and Bayesian Testing of a Precise Hypothesis, *Statistical Science, Vol. 12, No. 3 pp 133-160.*

Bernardo, J. A., and Smith, A. F. M., (1994), *Bayesian Theory,* John Wiley and Sons, New York.

Burnell, L., and Horvitz, E., (1995), Structure and Chance: Melding Logic and Probability for Software Debugging, *Communications of the ACM, Vol. 38, No. 3, pp31-41.*

Caves, C. M., and Fuchs, C. A., (1996), Quantum Information: How much Information in a State Vector, *quant-ph/9601025,* http://xxx.lanl.gov/archive/quant-ph.

Chang, K. C., (1994), Multi-Source Intelligence Correlation and Fusion with Bayesian Networks, *Technical Proceedings of the Seventh Joint Service Data Fusion Symposium,* The John Hopkins University Applied Physics Laboratory, Laurel Maryland.

Chang, K. C., Lui, J. and Zhou, J., (1996), Bayesian probabilistic inference for target recognition, *Proc. SPIE* **2755**, *158-165.*

Chang, K. C. and Fung, R., (1997), Target identification with Bayesian networks in a multiple hypothesis tracking system, *OPT. Eng., vol 36, no. 3, pp 684-691.*

Cheeseman, P. (1986), Probabilistic vs. Fuzzy Reasoning. In *Uncertainty in Artificial Intelligence,* (L. N. Kanal and J. F. Lemmer, ed.), North-Holland, Amsterdam.

Chuaqui, R., (1991), *Truth, Possibility and Probability, New Logical Foundations of Probability and Statistical Inferences.* North-Holland, Amsterdam

Cooper, G., (1990), Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks (Research Note), *Artif. Intell. Vol. 42, pp 393-405.*

Cover, T. M., and Thomas, J. A., (1991), *Elements of Information Theory,* John Wiley and Sons, New York.

Dagum, P., and Luby, M., (1993), Approximately Probabilistic Reasoning in Bayesian Belief Networks is NP-hard, *Artif. Intell. Vol. 60, pp 141-153.*

Discenza, J. H., Use of Monte Carlo Modelling and Bayes's Rule for Situation Assessment in the Ground Battle, *Report from Daniel H. Wagner Associates, Inc. 2 Eaton Street, Suit 500, Hampton VA 23669 SBIR*

Dubois, D., Prade, H. and Yager, R. R., eds. (1997), *Fuzzy Information Engineering: A guide tour of applications,* John Wiely and Sons INC., New York.

Durrant-Whyte, H. F., (1991), Sensor fusion: when more is better. In K. T. V. Grattan, ed., *Sensors: Technology, Systems and Applications* Adam Hilger, Bristol.

Efron, B., (1978), Controversies in the foundations of statistics, *Amer. Math. Monthly, Vol. 85, pp 231-246.*

Efron, B., (1998), R. A. Fisher in the 21st Century, *Statistical Science, Vol. 13, No. 2 pp 95-122.*

Endsley, M. R., (1995), Towards a theory of situation awareness in dynamic systems, *Human Factors, Vol. 37, pp 32-64.*

Fabian, I. and Lambert, D. A., (1998), First-Order Bayesian Reasoning, Preprint from authors.

Fine, T., (1973), *Theories of Probability.* Academic Press, New York.

De Finetti, B., (1972) *Probability Induction and Statistics,* John Wieley & Sons, London.

Fung, R., and Del Favero, B., (1995), Applying Bayesian Networks to Information Retrieval, *Communications of the ACM, Vol. 38, No. 3, pp42-48, 57.*

Heckerman, D., Breese, J. S., and Rommelse, (1995), K., Decision-Theoretic Troubleshooting, *Communications of the ACM, Vol. 38, No. 3, pp49-57.*

Heckerman, D. and Wellman, M. P.,(1995) Bayesian Networks, *Communications of the ACM, Vol. 38, No. 3, pp27-30.*

Jaynes, E. T., (1988) The relation of Bayesian and Maximum Entropy Methods, In (G. J. Erickson and C. R. Smith ed) *Maximum-Entropy and Bayesian Methods in Science and Engineering Vol. 1: Foundations,* Kluwer Academic Publishers, Doddrecht.

Jensen, F. V., (1996) *An Introduction to Bayesian Networks.* Springer, New york.

Kanal, L. N. and Lemmer, J. F., ed., (1986) *Uncertainty in Artificial Intelligence,* North-Holland, Amsterdam.

Kanal, L. N. and Lemmer, J. F., ed., (1988) *Uncertainty in Artificial Intelligence 2,* North-Holland, Amsterdam.

Lambert, D. A., (1989), Assets with ATTITUDE. *DSTO Technical Report to be published.*

Laskey, K. B., Stanford S., and Col. Stibio, B., (1994), Probabilistic Reasoning for Assessment of Enemy Intentions, Publication # 94-25, *Center of Excellence in C3I, George Mason University,* Fairfax, Virginia.

Levitt, T. S., Winter, C. L., Turner, C. J., Chestek, R. A., Ettinger, G. J. and Sayer, S. M., (1995), Bayesian inference-based fusion of radar imagery, military forces and tactical terrain models in the image exploitation system/balanced technology initiative, *Int. J. Human-Computer Studies vol 42, pp 667-686.*

Lindley, D. V., (1987), The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems, *Statistical Science, Vol. 2, No. 1 pp 17-24*

Lui, J. and Chang, K. C., (1996), Feature-based target recognition with a Bayesian network, *OPT. Eng., vol 35, no. 3, pp 701-707.*

Luo, R. C. and Kay, M. G., (1989), Multisensor Integration and Fusion in Intelligent Systems, *IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, No 5.*

Manka, M. and Nicholson, J. A., (1999), Bayesian Belief Networks: Definition, Construction and Applications to Special Forces Operations, *DSTO-DP-0692.*

Moffat, J., (1998), Representing the Command and Control Process in Simulation Models of Combat

Neapolitan, R. E.(1990), *Probabilistic Reasoning in Expert Systems,* John Wiley & Sons, Inc, New York.

Netica (1998), http://www.norsys.com/home.html.

Ottonello, C., Peri, M., Regazzoni, C., and Tesei, A., (1992), Integration of Multisensor Data for overcrowding estimation, *IEEE International Conference on Systems, Man and Cybernetics (Chicago, Ill.),* IEEE, New York Oct. 1992, pp 791-796

Pearl, J., (1988), *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc. San Francisco, California.

Pearl, J., (1998), Bayesian Networks, accessed through *http://singapore.cs.ucla.edu/csl_papers.html.*

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, (1993), R. G., Bayesian Analysis in Expert Systems, *Statistical Science, Vol. 8, No. 3 pp 219-283*

Spiegelhalter, D. J., and Lauritzen, S. L., (1990) Sequential Updating of Conditional Probabilities on Directed Graphical Structures, *NETWORKS, Vol. 20, pp 579-606.*

Staker, R. J., (1999), Information System Network Risk Analysis Using Bayesian Belief Networks, *DSTO-TR-0830*

Verma, T. S., (1986), Causal networks: Semantics and expressiveness. In *Proc. 4th Workshop on Uncertainty in AI, Minnaepolis, Minn. pp 352-59.*

Waltz, E. L., and Buede, D. M., (1986), Data Fusion and Decision Support for Command and Control, *IEEE Transactions on System Man and Cybernetics, Vol. SMC-16, No. 6 pp 865-879.*

Waltz, E. and Llinas, J., (1990) *Multisensor Data Fusion,* Artech House, Boston.

Wohl, J. G., (1981) Force Management Decision Requirements for Air Force Tactical Command and Control, *IEEE Transactions on System Man and Cybernetics, Vol. SMC-11, pp 618-639.*

Zsambok, C., and Klien, G., [Eds], (1997), *Naturalistic Decision Making,* Lawrence Erlbaum Associates, Hillsdale, NJ.

Appendix A: Bayesian Formalism

A.1 The Interpretations of Probability

Generally, different interpretations of fundamental scientific concepts rarely have any effect in the application of the corresponding scientific methods to real world problems. This is not the case when dealing with probability and statistics. Normally the differing interpretations do lead to identical results, however, it is not infrequent for them to yield mutually contradictory consequences [Chuaqui, 1991]. Here we briefly clarify these interpretational issues as it puts our endeavour in model building in proper perspective.

The interpretation of probability becomes important when we want to apply probabilistic theory to events occurring in the real world:

a. Given some event A what do we mean by the number $P(A)$ and therefore how to determine this number.

b. If we are given a number as the value of $P(A)$ what do we understand it as and therefore how can we use this number to make predictions about the occurrence or non occurrence of the event A.

Philosophers have long argued about the interpretation of the number representing the probability of an event, even when applied to mundane events that results from tossing a coin. Most workers adopt one of the following two interpretations [Fine, 1973; de Finetti, 1972]

A.1.1 The relative frequency or the objective interpretation

The relative frequency approach is based on the following definition:

The probability $P(A)$ of an event A is given by

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

Here we make a series of trials. The number n_A denotes the number of times the event A occurs in a series of n trials.

This interpretation is based on our general conviction that for the real world phenomenon in question, the ratio n_A/n does approach a limit when repeated trials are performed either sequentially or simultaneously and this limit remains the same when evaluated for any subsequence of the trials.

A.1.2 The subjective interpretation

In the subjective interpretation, the number *P(A)* is understood as a measure of our lack of *knowledge* or uncertainty about the occurrence or non occurrence of the event *A* in a single performance of the underlying experiment. We do not fully understand the phenomenon that gives rise to the event *A*, possibly because it is too complicated, and therefore are not sure if the event *A* will occur when a trial is performed. This ignorance is reflected in the number *P(A)*. This notion of probability is therefore personalistic, it reflects the degree of belief of a particular person at a particular time. The subjective interpretation, unlike the relative frequency interpretation, does not prescribe a general set of procedures by which to determine the number *P(A)*. This is rightly so, because in this interpretation probability is derived from a personal knowledge of the underlying phenomenon. Obtaining knowledge of some phenomenon is not a part of probabilistic considerations but possibly a part of physics or some other science. In most cases the aim is to obtain as much information about the underlying phenomenon as possible, then apply logical reasoning to arrive at a satisfactory number.

Different interpretations of the concept of probability lead to different statistical techniques. In fact the modern statistical thinking can be categorised along the lines of three competing schools: Bayesian, frequentist and Fisherian [Efron, 78]. Those who adopt the subjective interpretation of probability are Bayesian while the adherents of the objective point of view are frequentists. Most statisticians that work in testing of hypotheses based on statistical data are frequentists while most of those who work with problems in decision making are Bayesian. In many ways the Bayesian and the frequentist techniques stand at opposite poles from each other, with Fisherian techniques being somewhat of a compromise [Efron, 98; Berger et al., 97].

A.2 The Bayesian Formalism

The Bayesian formalism allows a decision maker to update his subjective belief when new facts are uncovered. The basic expression for conditional probability of an event *A* given an event *B* is given by the relation

$$P(A \mid B) = \frac{P(A, B)}{P(B)}.$$

Consider the joint probability *P(A,B)*, from the *commutative* property of logic it follows that the two propositions

(A,B) = A and B are both true

(B,A) = B and A are both true

are the same, and so they must have the same truth value and the same probability whatever our state of knowledge [Jaynes 1988]. Hence *P(A,B) = P(B,A)*. Applying this to the conditional probabilities we obtain the celebrated *Bayes' formula*

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

We can interpret this formula as follows:

> On the right hand side $P(B)$ represents our initial belief that B will materialise, this is often termed the *prior* belief. $P(A \mid B)$ represents the belief that some evidence A will be found once we have known that the event B has actually materialised, whereas $P(A)$ represents our belief that A will be found to be true under general circumstances.
>
> $P(B \mid A)$ represents our revised belief. The belief that B will occur or has occurred after we come to know that some evidence A supporting or denying B has been uncovered. This is called the *posterior* belief in B.

Bayes' formula therefore gives us a definite procedure to update our belief. Our *prior* belief is updated to the *posterior* belief, after supporting evidence has been uncovered. What is more, anyone who reasons in a way that conflicts with Bayes' formula would then be violating a rather elementary principle of logic. This is no doubt a salutary thought.

Appendix B: Bayesian Networks

A complete and rigorous definition of Bayesian networks can be found in some of the literature cited [Pearl 1988, Neapolitan 1990, Jensen 1996]. Here we will develop the concept with just enough rigour and detail that will enable us to apply these networks to C3I problems. For convenience we shall use slightly different but not contradictory notations in this appendix. Random variables are again denoted by X, Y, Z, ..., etc.- or by subscripted letters X_1, X_2, X_3, ..., etc. The values taken by these variables are denoted by lower case letters x, y, z, ..., etc.- or correspondingly by x_1, x_2, x_3, ..., etc. If all the values that are possible for X lie in the domain D_x then x represents any element in D_x. Boldfaced capital letters **X, Y, Z,** etc. represent sets of random variables and assignment of values to these sets, also called *configuration*, is done through bolbfaced lowercase letters **x, y, z,** etc. For example if **Z** stands for the set $\{X, Y\}$ then **z** represents the configuration $\{x, y\}$. $P(\mathbf{Z}=\mathbf{z})$ is equivalent to the joint distribution $P(X=x, Y=y)$. We will use the short form $P(\mathbf{z})$ and $P(x,y)$ to denote these probabilities instead.

Let us pave the way by an example [Pearl 1998].

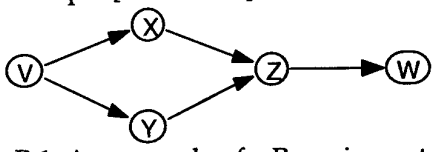

Figure B.1. An example of a Bayesian network

Figure B.1 is a simple Bayesian network. It describes the causal relationships among the season of the year (V), whether rain falls (X) during season, whether the sprinkler is on (Y) during that season, whether the pavement would get wet(Z), and whether the pavement would be slippery (W). A wet pavement will cause a slippery pavement; the arrow from Z to W represents this causal relationship. There are no direct connections between V and W, and this captures our understanding that the influence of seasonal variations on the slipperiness is mediated by other conditions. With this example in mind let us proceed with a semi formal definition.

B.1 Causal Networks

A Causal network is a *directed acyclic graph (DAG)*, like one shown in figure B.1, in which the nodes represent the random variables of a probabilistic model and the directed links represent *informational* or *causal* dependencies among the variables. A parent in the network is deemed to be a cause influencing its children. The strength of the causal dependency is represented by conditional probabilities that are attached to each cluster of parent→child links in the network.

In the case of Fig B.1 the values $P(z \mid x,y)$ - numbers giving conditional probabilities for all possible states of X, Y, and Z - quantify reasoning in the direction along the links from $\{X, Y\}$ to Z. If we obtain evidence regarding cause $\{X, Y\}$ then the conditional probabilities allow us to calculate the effect on Z.

Reasoning in the direction opposite to the link is accomplished through Baye's rule, as will be shown later; evidence pertaining to the effect prompts an updating of belief regarding the cause.

B.2 Conditional Independency

Let $\boldsymbol{U}$ denote a finite set of discrete random variables. Let $P(\cdot)$ be a joint probability function over the variables in $\boldsymbol{U}$ and let $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, stand for any three subsets of variables in $\boldsymbol{U}$. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be *conditionally independent given* $\boldsymbol{Z}$ if

$$P(\boldsymbol{x} \mid \boldsymbol{y},\boldsymbol{z}) = P(\boldsymbol{x} \mid \boldsymbol{z}) \text{ whenever } P(\boldsymbol{y},\boldsymbol{z}) > 0.$$

We shall use the notation $I(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{Y})$ to denote the conditional independence of $\boldsymbol{X}$ and $\boldsymbol{Y}$ given $\boldsymbol{Z}$

B.2.1 Independency and separation

Let us now examine the correspondence between the topology of a causal network and the independencies portrayed by it

Serial Connections

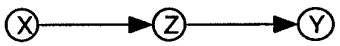

Figure B.2. Causal network with serial connections.

In fig B.2 evidence regarding X will influence the knowledge about Z this in turn will influence the knowledge about Y. Similarly if we obtain evidence about the effect Y then our belief in its cause Z will be updated this in turn will affect our belief in the cause of Z which is X. However if Z is instantiated then knowing the state of Z is enough to determine the probabilities of various states of Y, because of the conditional probability attached to the link $Z \rightarrow Y$. Evidence regarding X will not be able to influence knowledge about Y any more. Similarly evidence about Y will be blocked by the instantiated Z and will not affect knowledge about X. We say that in the serial connection depicted in figure B.2 X and Y are *separated* by Z.

Diverging Connection

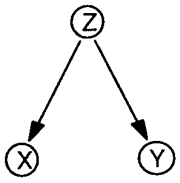

Figure B.3. Causal network with diverging connections.

In figure B.3 if we obtain evidence regarding the effect X then our knowledge about its cause Z will change. This will in turn influence our belief regarding the other effect of Z which is Y. However if Z is instantiated then knowing the state of Z is enough to determine the probabilities of various states of Y, because of the conditional probability attached to the link $Z \rightarrow Y$ and similar reasoning applies for X. Hence the effects X and Y become independent when Z is instantiated. We describe the situation by saying that in the diverging connection depicted in figure B.3, X and Y are *separated* by Z.

Converging Connection

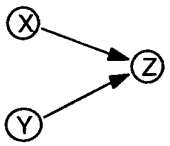

Fig B.4. Causal network with converging connections.

Reasoning about converging connections needs some care. Let us explain it through an example. In figure B.4, X represents the states *head* or *tail* of *Coin1* and Y represents similar states of *Coin2*. Z represents a bell which rings if in an experiment of tossing the coins both are found to have the same outcome. In general X and Y are independent. However, if the state of the bell Z gets instantiated then X and Y become dependent. This is called *induced or conditional dependency*.

The three cases described above cover all the ways in which influences can travel in a causal network. The separations and dependencies described capture all the ways in which the nodes can become independent or dependent. We capture this fact rigorously in the following definition:

B.2.2 d-Separation

> If X, Y and Z are three disjoint subsets of nodes in a *DAG*, then Z is said to *d-separate* X from Y, denoted $<X|Z|Y>$, if along every path between a node in X and a node in Y there is a node W satisfying one of the following two conditions: (1) W has converging arrows and none of W or its descendants are in Z, or (2) W does not have converging arrows and W is in Z.

If a path satisfies the condition above it is said to be blocked; otherwise it is said to be activated by Z.

Example

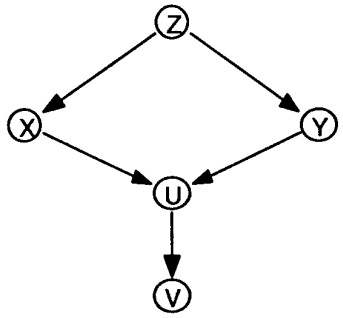

Figure B.5. A network illustrating *d-separation.*

Here the node Z *d-separates* the node X from the node Y. The path $X \leftarrow Z \rightarrow Y$ between X and Y is blocked by Z. The path $X \rightarrow U \leftarrow Y$, between X and Y is also blocked because U and its descendants are outside Z. However, X and Y are not *d-separated* by the set of nodes $\{Z, U\}$, because the path $X \rightarrow U \leftarrow Y$, between X and Y is now active; knowing U will make X and Y dependent.

B.3 Bayesian Network for a Probabilistic Model

The following stipulations define a *Bayesian Network* for a probabilistic model

- Consider a probabilistic model M consisting of a set $\boldsymbol{U}$ of discrete random variables and a joint distribution $P(\cdot)$ defined over these variables. Let D be a *DAG* whose set of vertices has a one to one correspondence with $\boldsymbol{U}$.
- D is said to be an *I-map* of the probabilistic model M if every *d-separation* condition displayed in D corresponds to a valid conditional independence relationship in M, i.e., for every three disjoint sets of vertices X, Y and Z we have

$$<X|Z|Y> \Rightarrow I(X,Z,Y)$$

- D is a *minimal I-map* of M if non of its arrows can be deleted without destroying its *I-mapness*.
- A *Bayesian network* of the probabilistic model is by definition a *DAG* D which is a *minimal I-map* of M

B.3.1 Constructing a Bayesian Network for a Probabilistic Model

The following questions arise immediately

- Given a probabilistic model, how can we construct a Bayesisn network?
- How can this Bayesian network help us in making probabilistic inference?

Construction of Bayesian networks is facilitated by a key theorem proved by Verma [Verma 1986]. Instead of discussing this theorem, let us lay down the steps that explain the construction of a Bayesian Network [Heckerman and Wellman 1995].

1. Consider a probabilistic model consisting of a set of discrete random variables $\boldsymbol{U}$ together with a joint probability distribution $P(\cdot)$ defined over these variables.
2. Let $X_1, X_2, \ldots, X_n$, be any ordering d of the random variables in $\boldsymbol{U}$. In this ordering let us explicitly denote the joint probability distribution as $P(x_1, x_2, \ldots, x_n)$.
3. For any X_i take the minimal subset $\Pi_i \subseteq \{ X_1, X_2, \ldots, X_{i-1}\}$ such that Π_i renders X_i, and $\{X_1, X_2, \ldots, X_{i-1}\}-\Pi_i$ conditionally independent. That is $P(x_i \mid x_1, \ldots, x_{i-1}) = P(x_i \mid \Pi_i)$
4. Create a *DAG* by designating the elements in the subset Π_i as parents of X_i.
5. Verma's theorem asserts that such a *DAG* would be a Bayesian network of the probabilistic model.

Hence given a probabilistic distribution choose an ordering $X_1, X_2, \ldots, X_n$, of $\boldsymbol{U}$. Start by designating X_1 as a root node and assign it the marginal probability $P(x_1)$, dictated by the joint distribution $P(x_1, x_2, \ldots, x_n)$. Next form the node X_2, if X_1 is a cause for X_2 i.e. if $P(x_2) \neq P(x_2 \mid x_1)$ then establish a directed edge from X_1 to X_2 and quantify this link by the conditional probability $P(x_2 \mid x_1)$. If X_1 does not influence X_2 then make X_2 a root node and assign it the marginal probability $P(x_2)$. This process of creating nodes and linking them continues in the order chosen. At the ith stage, we form the node X_i, following the dictates of the *step-3* above we determine the minimal subset $\Pi_i \subseteq \{ X_1, X_2, \ldots, X_{i-1}\}$. The nodes in Π_i are parents of X_i, we represent this by drawing the appropriate parent to child links and quantify this bunch of links by the conditional probability $P(x_i \mid \Pi_i)$.

The structure of the Bayesian network constructed above depends upon the node ordering d used in the construction. It can be shown that any other ordering d'

consistent with the direction of arrows of an already constructed network has the same network topology [Pearl 1988]. Hence, once the network is constructed, the original order can be forgotten. What does matter are the local orderings displayed by the network topology.

Bayesian networks can be viewed as graphical inference engines for deriving probabilistic inferences. Although it can be used, for numerical calculations of probabilities more efficiently, its main attraction lies in its ability to bring forth the logical dependencies inherent in the probabilistic model. For example, new independence relationships from those used in the construction of the network can be easily deduced by just inspecting the topology of the network.

However, the most interesting aspect of the Bayesian network formalism, lies not in its ability to reason about probabilistic models more efficiently, but in its ability to create probabilistic models *ab initio*. We discuss this in the next section.

B.4 Constructing Probabilistic Models for C3I problems

In this section, given a C3I problem, we show how to construct a Bayesian network for representing the uncertainties in this problem. A probabilistic model then emerges from the network.

In the light of the topics discussed in the previous sections, a process of model development would comprise the following three stages:

1. *Qualitative stage:* Here the general relationships between the variables of interest, in terms of the relevance of one variable to others are taken into account. This would result in a graphical representation capturing the conditional dependencies in a qualitative i.e. non-numerical fashion.
2. *Quantitative stage:* The links in the graphical representation are then assigned numbers representing conditional probabilities. This allows computation of probabilities that are of interest.
3. *Modification stage:* Here lessons learnt are incorporated into the model and any discrepancies with empirical data or other conflicts are either explained or removed.

B.4.1 Qualitative Stage

Bayesian networks provide a direct model of the real world environment rather than a model of the reasoning process as is done in many knowledge representation schemes e.g. neural networks etc. When building a probabilistic model of a real world C3I situation, one starts by identifying all the components of interest and associates random variables to them. These random variables then assume states corresponding

to the various possible states of the respective components. The knowledge about the system and intuitive understanding of various dependencies are then used to construct the causal structure of the C3I system. Here the graphical representation becomes very handy. It permits users to express directly the fundamental, qualitative relationships of direct influence.

A process for obtaining an initial graph can be along the following lines. In the first stage we examine each variable to find out if it is a root cause or is directly influenced by any other variables. All root causes are then assigned a node each. Let us call these nodes *level-1* nodes. We then find out all the variables that are directly influenced by the variables in the *level-1* nodes and no other variables. By assigning nodes to these variables we obtain a set of *level- 2* nodes. A given node at *level-2* has as its parents all those nodes in *level-1* that directly influence this particular node. A set of parent→child links are then drawn which now serve as edges of the graph. In the *(i+1)th* stage we identify all the variables that are directly influenced by variables in the preceding *i levels* and no other variables, and add the *level-(i+1)* nodes. The parents of these nodes are identified and the corresponding parent→child links are added. This hierarchical process continues until all the variables have a place in the graph and all parent→child links are accounted for by edges of the graph.

It must be stressed here that the whole process is subjective, because the parents are identified through the subjective judgement of the individual constructing the graph. This procedure is, however, consistent because in the Bayesian network formalism, for any node, once the direct influences on it are known, all other potential influences are irrelevant as far as constructing the network is concerned. The network then augments these with derived relationships of indirect influences in a consistent manner. To this end it must be mentioned that we have made a tacit assumption- the subjective judgements of causality do not lead to a cycle of causality.

B.4.2 Quantitative Stage

Once an initial network is in place, the attention shifts to providing the conditional probabilities to specify the strengths of the direct influences discussed above. These probabilities are again essentially subjective probabilities. One depends upon one's personal judgement, or consults users and experts, to arrive at subjective estimations of the conditional probabilities.

To provide an ordering, we start with the variables at *level-1* and serialise them as X_1, X_2, ..., in some order. After all the variables in *level-1* are exhausted we take up the variables at *level-2* and proceed down the hierarchy until we get a complete ordering of the form $X_1, X_2, \ldots, X_n$, say. Given any node X_i, let Π_i denote its parents. The subjective probabilities are only required locally to be assigned to the bunch of $\Pi_i \rightarrow X_i$ (parent→child) links as conditional probabilities $P(x_i \mid \Pi_i)$. For consistency with the axioms of probability one has to make sure that these assessments satisfy the relation

$$\sum_{x_i} P(x_i \mid \Pi_i) = 1 \text{ with } 0 \le P(x_i \mid \Pi_i) \le 1$$

The network formalism then provides the joint probability distribution over all the variables through the relation:

$$P(x_1,...,x_n) = \prod_i P(x_i \mid \Pi_i)$$

These joint probabilities then provide quantitative assessments of the problem domain that has been modelled. By this we mean probabilistic inferences of the form $P(S_1 \mid S_2)$ where S_1 and S_2 each represent a conjunction of a number of instantiated variables. Although a number of efficient algorithms exist to calculate such probabilities, the problem is essentially NP-hard [Cooper, 1990; Dagum and Luby, 1993]

It is the subjective nature of the conditional probabilities $P(x_i \mid \Pi_i)$, provided by the user or the expert, that brings the model closer to reality as humans perceive. Essentially what is required is to assess that an event would occur given a particular state of the environment that directly influences the event in question. These kinds of assessments are natural to human experts and capture the essence of the empirical knowledge acquired through experience. In other words, the resulting network captures the *belief* of the agent who constructs the network. Bayesian networks are therefore often referred to as *Bayesian Belief Networks (BBN).*

B.4.3 Modification Stage

As making probabilistic inferences with the network proceeds, the initial network may need some modifications. Some links may appear superfluous and some initial beliefs may have to be changed. For this purpose *learning techniques* have been developed for systematic updating of the conditional probabilities, as well as the structure of the network, so as to match empirical data [Spiegelhalter and Lauritzen, 1990].

Furthermore, the situation being modelled may also undergo change requiring a corresponding change in the network. Such changes are also easily implemented. In fact to quote Pearl [Pearl 1998]:
"The most distinctive feature of Bayesian networks, stemming largely from their causal organisation, is their ability to represent and respond to changing configurations. Any local reconfiguration of the mechanisms in the environment can be translated, with only minor modification, into an isomorphic reconfiguration of the network topology."

For example, if we need to add a new node X_{new} say, then all we need to do is to identify the set of nodes Π that have direct influence on X_{new} and provide a quantification of the local links $\Pi \rightarrow X_{new}$. We need not worry about the effects X_{new} would have on nodes outside Π. The network formalism guarantees this local change to be consistent with the global topology.

B.5 Propagation of Evidence in a Bayesian network

Suppose we have built a Bayesian network representing the uncertainties in some problem we are interested in. If evidence arrives indicating the states of some random variables, how would we update our knowledge regarding the states of other random variables. We discuss this point here. Again we will not be able to discuss evidence propagation in all its generalities, the interested reader may consult literature on the theory of Bayesian networks cited before. All we intend to do is give enough details so the applications we discuss in this document become transparent.

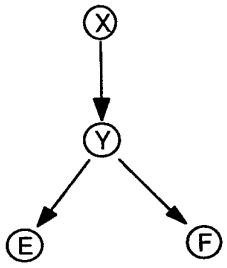

Figure B.6. A Bayesian network illustrating evidence propagation.

Consider the Bayesian network depicted above. The network structure provides the conditional probabilities $P(y \mid x)$, $P(e \mid y)$, and $P(f \mid y)$. We consider two modes of propagation of evidence:

Propagation along the Links:

If evidence arrives that X has assumed the value x_1, then the probability distribution over the states of Y are given by $P(y \mid x_1)$ which can be directly calculated from the conditional probabilities. Similarly $P(e \mid x_1) = \sum_y P(e \mid y)P(y \mid x_1)$ is also calculated by the network conditional probabilities.

In general when we fix the prior probability distribution over the node X, depending upon the pre-existing knowledge, this distribution propagates along the links to fix the prior probability distribution over all other nodes. In other words propagation along the links fixes the prior distributions.

Propagation against the link

If evidence arrives that E has assumed the value e_1 and F has assumed the value f_1, then these evidence update the probability distribution over the states of X according to the dictates of Baye's rule:

$$P(x \mid e_1, f_1) = \frac{\sum_y P(f_1 \mid y)P(e_1 \mid y)P(y \mid x)P(x)}{\sum_{yx} P(f_1 \mid y)P(e_1 \mid y)P(y \mid x)P(x)}$$

Here $P(x \mid e_1,f_1)$ are the posterior probabilities whereas $P(x)$ occurring in the right hand side are the prior probabilities

Hence propagation against the links updates the prior distributions.

Generally Bayesian networks are constructed so that the root nodes, like X above, represent variables that are not directly observable. Variables that are observable often occupy the lower levels of a network. When observational evidence arrives these are channelled through the intermediate nodes to the root nodes. Probability distributions are updated by the dictates of Baye's rule as shown above. Directly observable variables are called *information variables*. Variables that are not directly observable and the knowledge about whose states are inferred from the evidence gathered regarding information variables are called *hypothesis variables*. In the above X is a hypothesis variable which, say, represents the disease suffered by a patient. E and F represent symptoms, these are directly observable. Y can be some medical condition. In general therefore evidence propagates from the information variables to the hypothesis variables in the direction against the links.

Appendix C: Situation Assessment; A Case Study

In this appendix a particular evolution of situation, for the scenario constructed in section 2.5, is considered. We demonstrate the use of Bayesian networks for updating our belief in the hypothesis variable *enemy intentions.*

After receiving intelligence about the trade pact between the enemy and the neutral country, HMAS ANZAC is tasked to put the *designated zone* under constant surveillance.

Stage 1. At this initial stage *enemy intention* is more likely to be *passive.* Figure C.1 shows the network with such a prior distribution. A neutral tanker is detected at a subsequent point of time near the *zone boundary* moving *slowly* while *communicating with base.* Figure C.2 utilises this information to update belief in hypothesis. Following the recommended course of action HMAS ANZAC despatches the patrol boat which turns the tanker back to its port of origin.

Stage 2. As a result of the action taken by HMAS ANZAC *enemy intention* is very likely to become defensive – Figure C.3 reflects this apprehension. An Orange *FCPB* approaches the restricted *zone* to conduct reconnaissance moving *slowly parallel* to the zone boundary. Figure C.4 uses these data to update belief and suggest Blue COA. HMAS ANZAC sends the F111 for attack at which point the patrol boat retreats out of the designated area, the F111 is withdrawn.

Stage 3. The enemy now being aware of the intention and possibly the size of the Blue forces dispatches a *communication ship* which takes up position near the boundary of the *designated zone* and commences *EW* activity. This is detected by HMAS ANZAC and the JOR is advised of the situation. Figures C.5 and C.6 depict the prior and posterior knowledge.

Stage 4. At this point of development the enemy intention is more likely to turn offensive. An Orange force frigate enters the restricted zone and takes up position to defend the supply rout. The Orange force frigate is detected and JOR informed. Figures C.7 and C.8 depict the prior and posterior knowledge about the situation. Although the recommended course of action favours launching a *minor offensive,* JOR orders HMAS ANZAC to mount a full scale attack the orange frigate coordinating the attack with F111 and pursue the attack until there is a 90% certainty that the frigate is crippled.

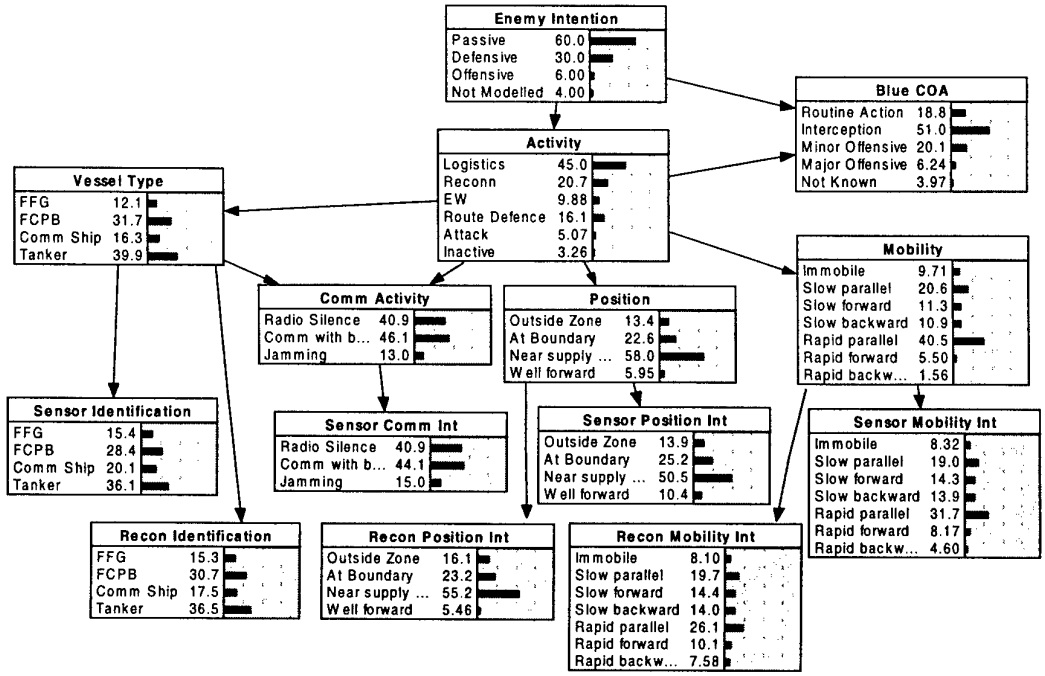

Figure C.1. Stage 1 prior probability distributions.

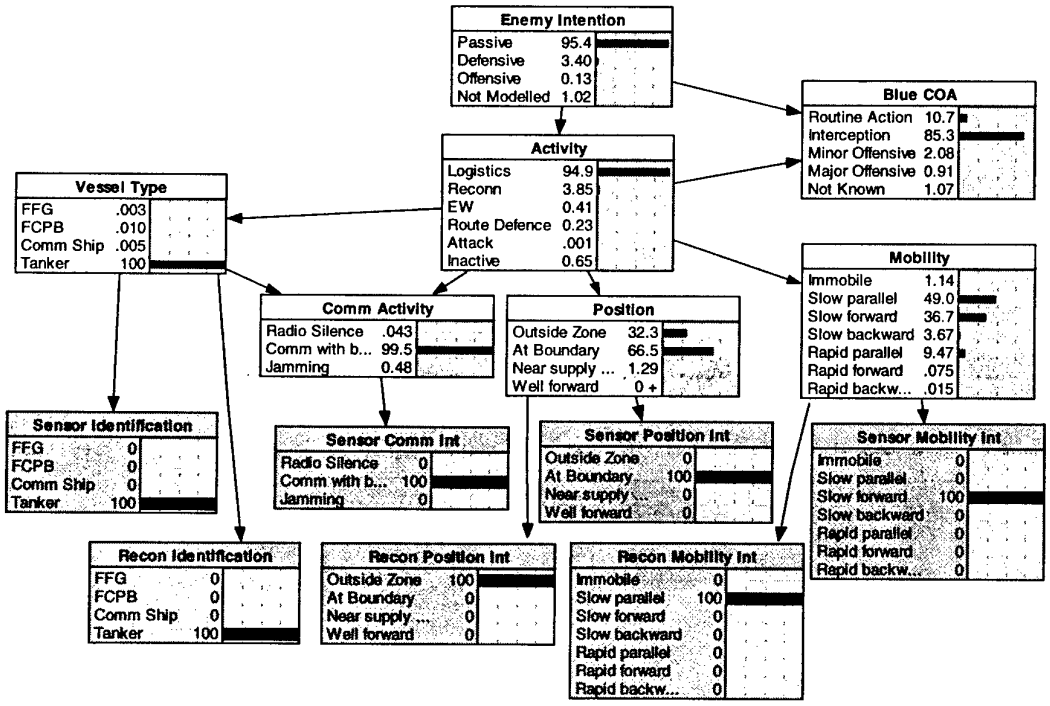

Figure C.2. Stage 1 posterior probability distributions.

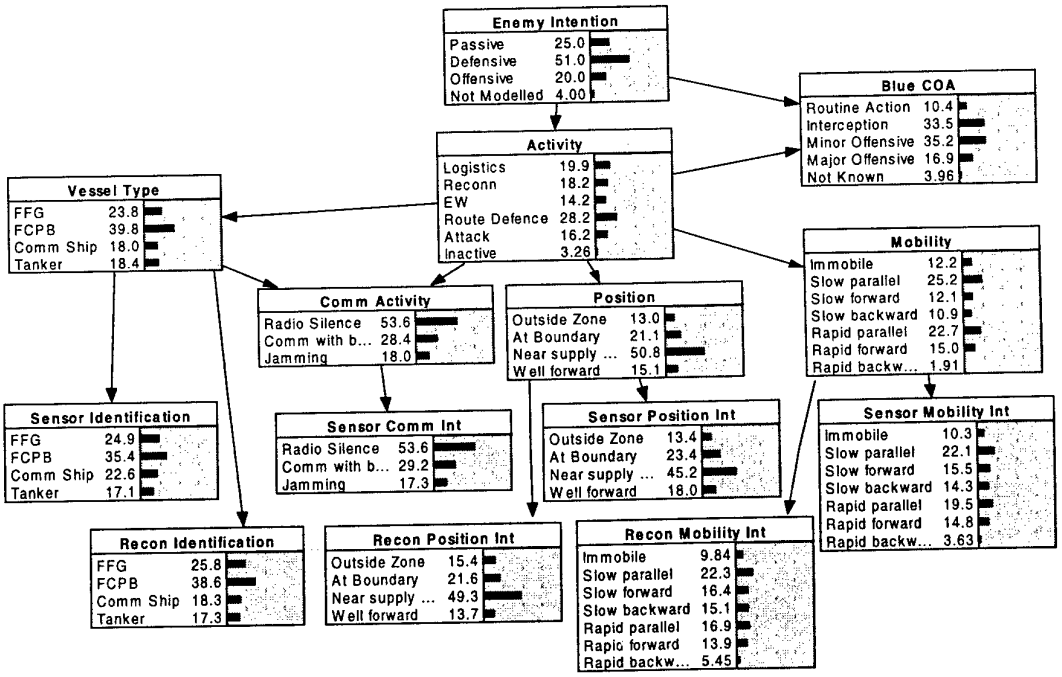

Figure C.3. Stage 2 prior probability distributions.

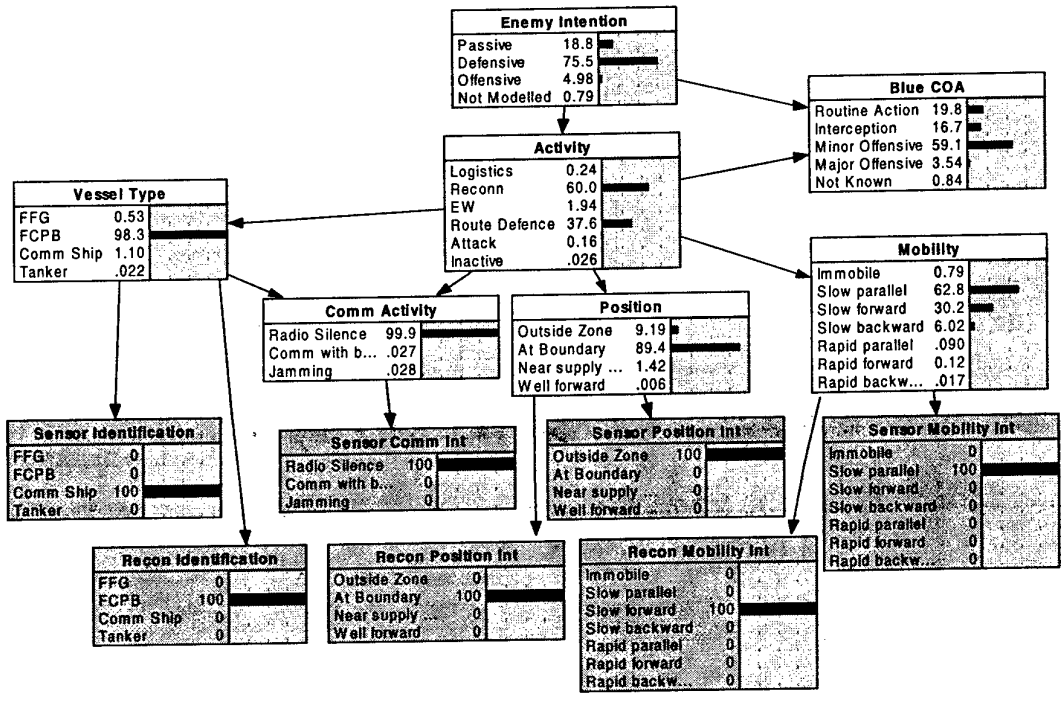

Figure C.4. Stage 2 posterior probability distributions.

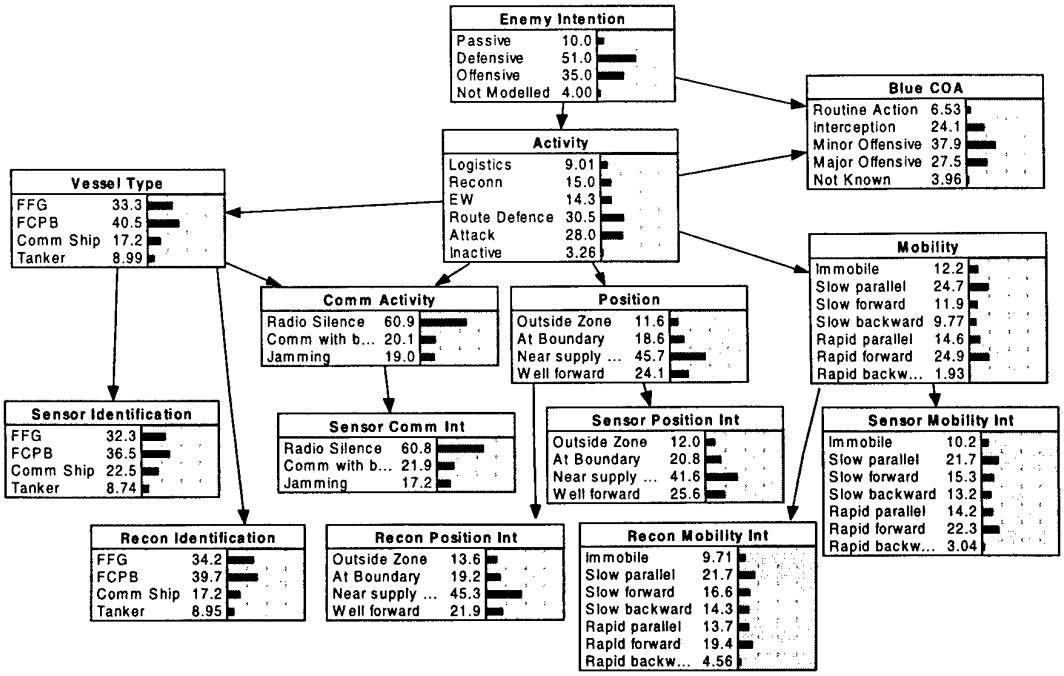

Figure C.5. Stage 3 proir probability distributions.

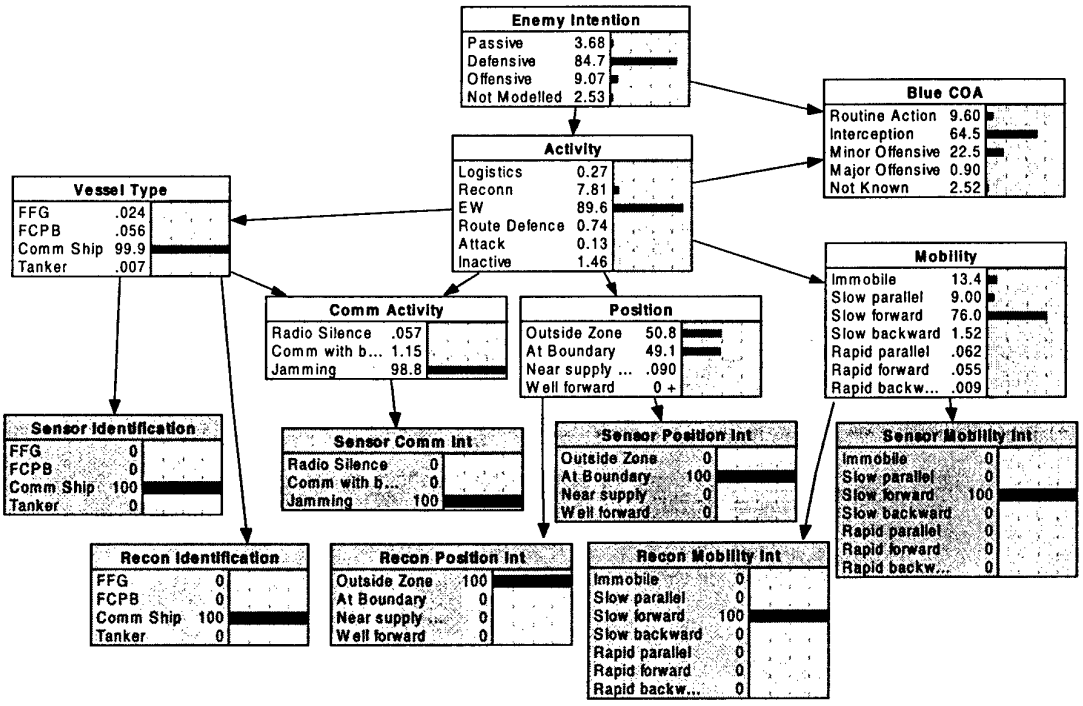

Figure C.6. Stage 3 posterior probability distributions.

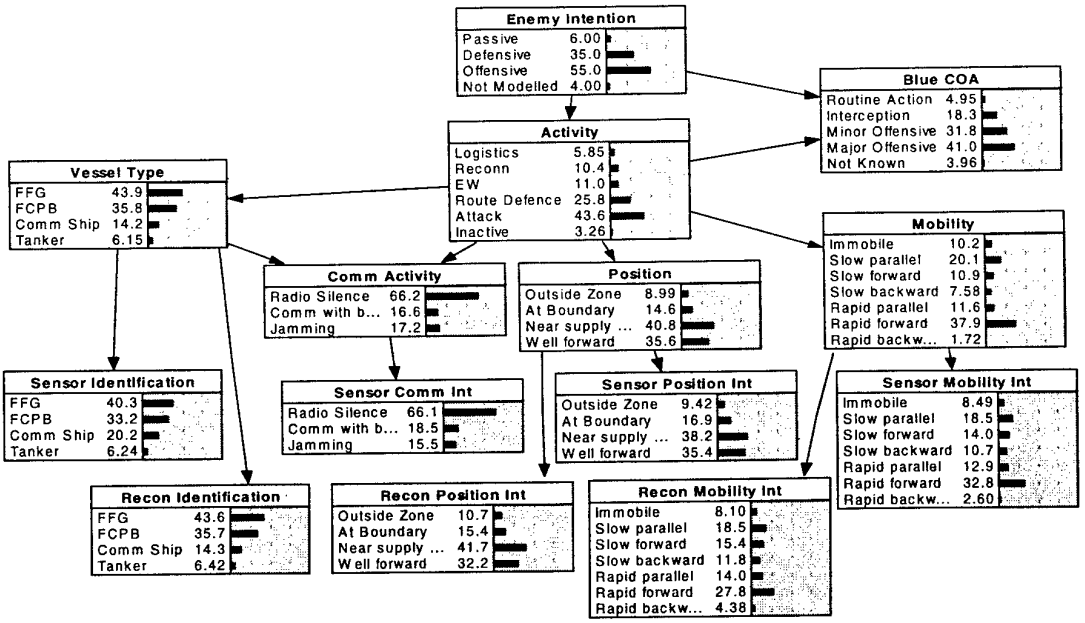

Figure C.7. Stage 4 prior probability distributions.

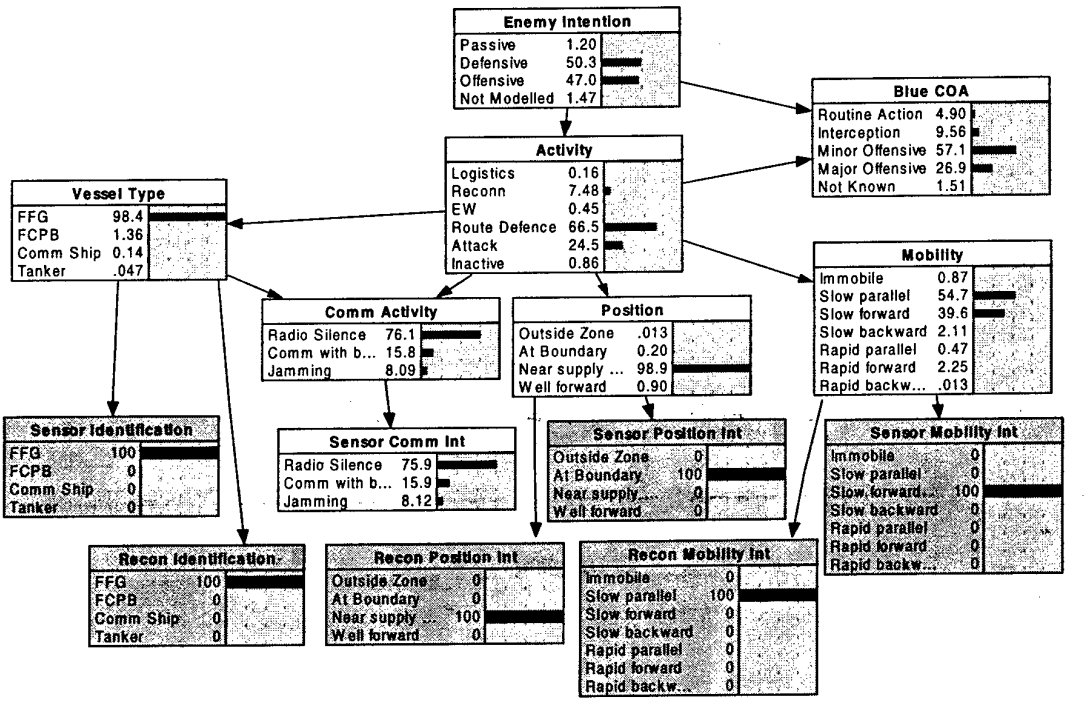

Figure C.8. Stage 4 posterior probability distributions.

Appendix D: Calculating Effectiveness

In this appendix we calculate some of the effectiveness measures mentioned in section 3. All computations are with regard to the Bayesian network for situation assessment as depicted in Figure 2.2, except that we have simplified the network by removing the reconnaissance nodes. This does not affect the generality of the computing procedures, it just makes the computations less cumbersome.

D.1 Evaluating the effect of evidence

Consider the Bayesian network for the situation assessment problem. Let us choose *Activity* as the target hypothesis variable. We want to calculate the reduction in uncertainty in *Activity* when evidence is obtained regarding the information variables *Position* and *Mobility*. Since *Activity* is not a root node, we modify the network as shown in Figure D.1. Here the node *Activity* has been disconnected from its parent node *Enemy Intention* and becomes a root node consequently. All the states in *Activity* have the same prior probability giving it the maximum prior uncertainty:

$$H(Activity) = \log 6 = 2.5849 \text{ bits}.$$

Figure D.2 shows the situation where we have incorporated the following evidence:

Sensor Position Int (SPI) := Near Supply Route

Sensor Mobility Int (SMI) := Slow Parallel

The conditional probabilities can now be read off from the *Activity* node. This immediately leads to the conditional entropy

$$\begin{aligned} &H(Activity \mid SPI = \text{Near supply route}, SMI = \text{Slow parallel}) \\ &= -0.1024\log 0.102 - 0.276\log 0.276 - 0.0224\log 0.0224 - 0.484\log 0.484 \\ &\quad - 0.0365\log 0.0365 - 0.0784\log 0.0784 \\ &= 1.9410 \end{aligned}$$

The uncertainty reducing capacity of the above evidence is $2.5849 - 1.9410 = 0.6439$ *bits*

D.2 Gain in Belief Updating

Consider a stage during situation assessment when previous evidence strongly suggests that enemy intention is defensive. This situation is depicted in Figure D.3 where the Bayesian network is initialised with a prior probability distribution P_1 indicating defensive enemy intentions. Figure D.4 depicts the updated situation after the latest batch of sensor evidence have been used to update the prior belief. The updated probability distribution over enemy intention is P_2.

x	*Passive*	*Defensive*	*Offensive*	*Not Modelled*
$P_1(x)$	*0.03*	*0.90*	*0.06*	*0.01*
$P_2(x)$	*0.0032*	*0.425*	*0.565*	*0.0062*

Whereas P_1 was peaked P_2 is flat. The loss of peak indicates a considerable change in the nature of the situation. The enemy has changed its intentions significantly. It would be unwise to stick to P_1 although that indicated more definitely what the enemy proposed to do. Although P_2 incorporates more uncertainty than P_1 it is a better reflection of the latest observations. The efficiency gained due to updating information is given by $D(P_2\|P_1)$ which can be calculated using the formula

$$D(P_2\|P_1) = -\sum_x P_2(x)\log P_1(x) + \sum_x P_2(x)\log P_2(x) = 2.41953 - 1.06201 = 1.35752 \text{ bits}$$

D.3 Effectiveness of the Position Detection Sensor

Consider the situation assessment network again. We denote by

X: Enemy Intention

Y: Position

Z: Sensor Position Int

The effectiveness of the position detection sensor is then given by $I(X;Y) - I(X;Z)$. To calculate this we initialise the network with a prior probability which gives maximum entropy to X as shown in figure D.5. We get

P(X):

x	*Passive*	*Defensive*	*Offensive*	*Not Modelled*
P(x)	*0.25*	*0.25*	*0.25*	*0.25*

P(Y)

y	*Outside Zone*	*At Boundary*	*Near Supply Route*	*Well Forward*
P(y)	*0.108*	*0.27*	*0.453*	*0.170*

P(Z)

z	*Outside Zone*	*At Boundary*	*Near Supply Route*	*Well Forward*
P(z)	*0.139*	*0.286*	*0.375*	*0.201*

To obtain $P(X|Y)$ and $P(X|Z)$ we successively instantiate Y and Z to their various values and allow the network to calculate the conditional probabilities. Figure D.6 depicts the case where we have instantiated Y= *Outside Zone*. We get

$P(x \mid y)$:

	x = Passive	x = Defensive	x = Offensive	x =Not Modelled
y = Out Zone	0.284	0.406	0.221	0.089
y = At Boundary	0.221	0.232	0.0868	0.47
y = Near Supply	0.355	0.293	0.172	0.180
y = Well Forward	0.0114	0.0648	0.736	0.188

$P(x \mid z)$

	x = Passive	x = Defensive	x = Offensive	x =Not Modelled
z = Out Zone	0.258	0.339	0.174	0.229
z = At Boundary	0.248	0.258	0.119	0.375
z = Near Supply	0.316	0.270	0.218	0.196
z = Well Forward	0.124	0.140	0.548	0.188

The above data yields
$H(X) = 2$, $H(X \mid Y) = 1.744433$, $H(X \mid Z) = 1.897761$ *bits*
Hence
$I(X ; Y) = H(X) - H(X \mid Y) = 0.25567$ *bits* and
$I(X ; Z) = H(X) - H(X \mid Z) = 0.102239$ *bits*
Hence the uncertainty reducing capacity of the position detection sensor defers from that of an ideal sensor by an amount:
$I(X;Y) - I(X;Z) = 0.15343$ *bits.*

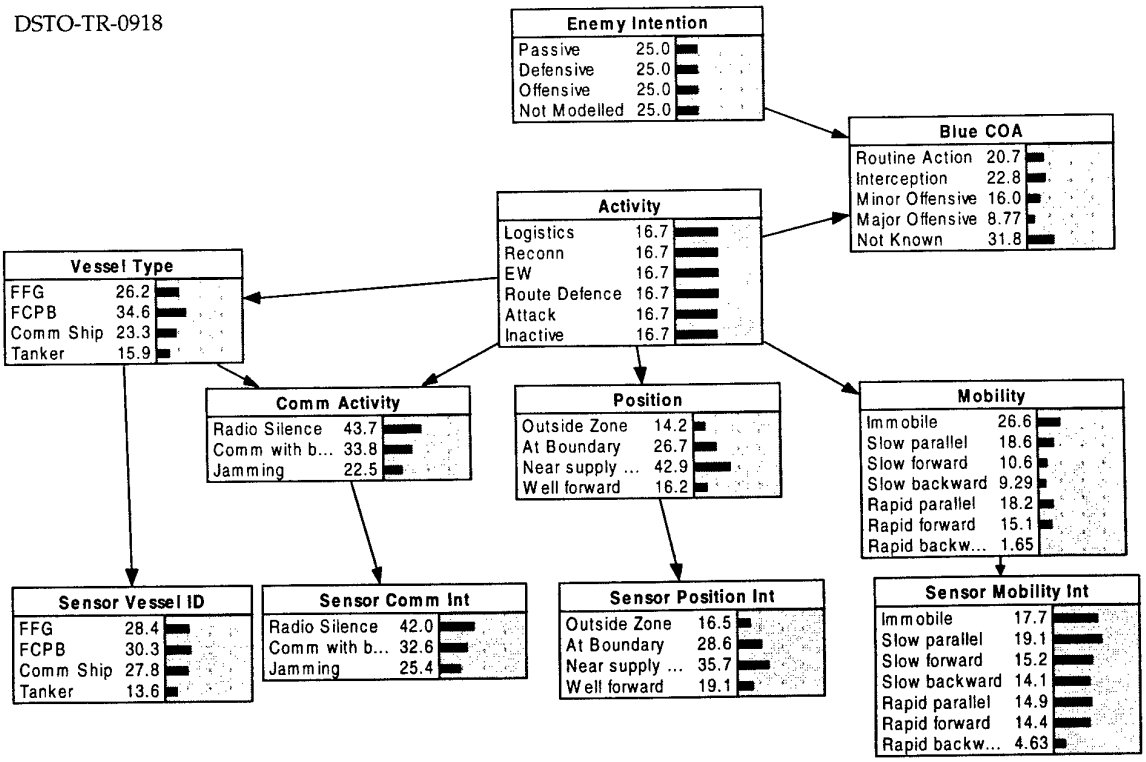

Figure D.1. The node *Activity* has been disconnected from its parent node *Enemy Intention* and as a consequence becomes a root node.

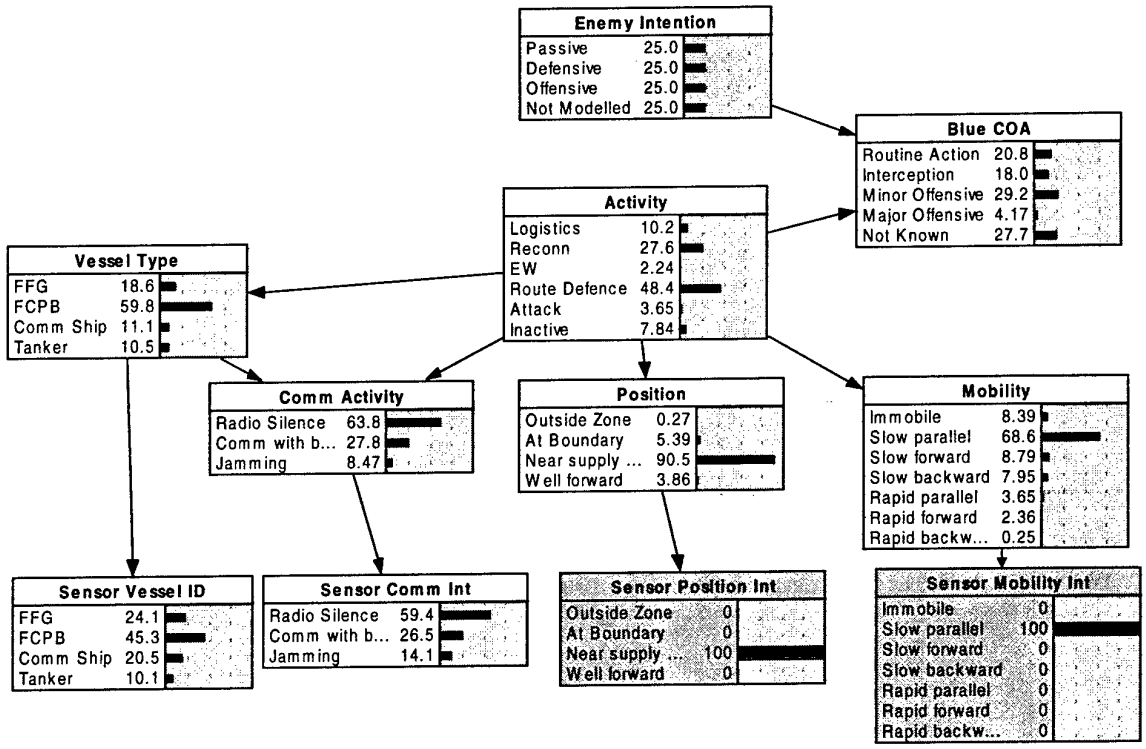

Figure D.2. Updated situation for evidence *Sensor Position Int (SPI) := Near Supply Route Sensor Mobility Int (SMI) := Slow Parallel*

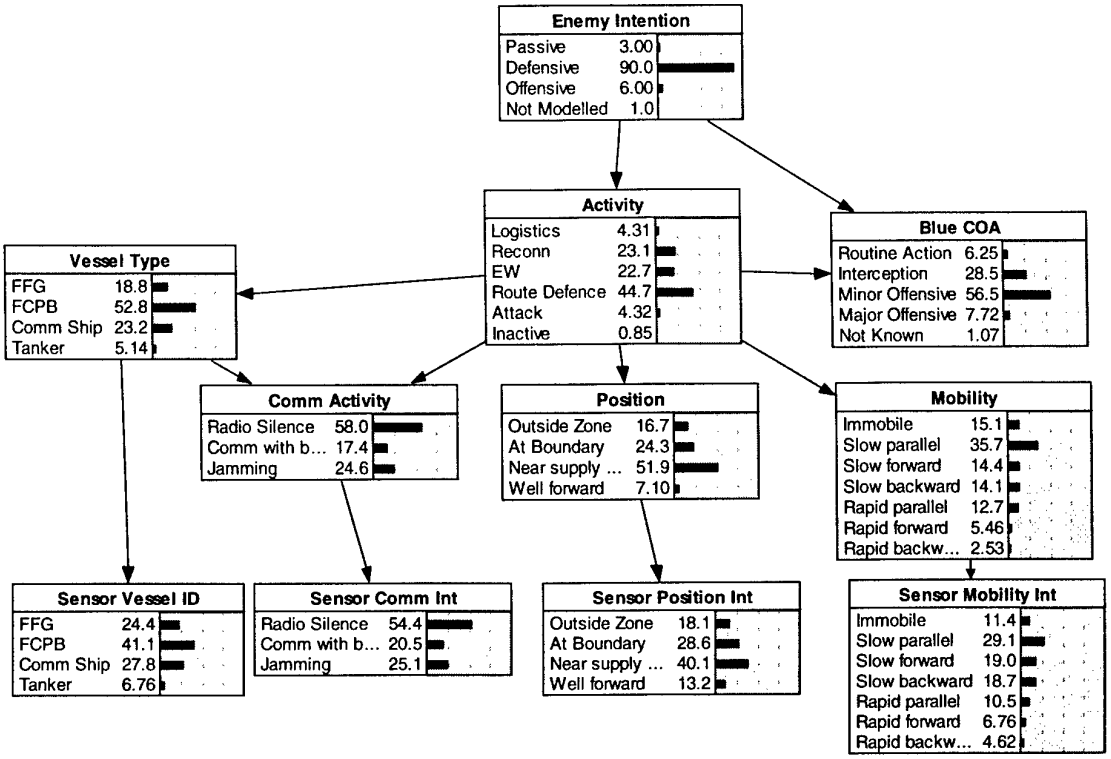

Figure D.3. Bayesian network initialised with a prior probability distribution P_1

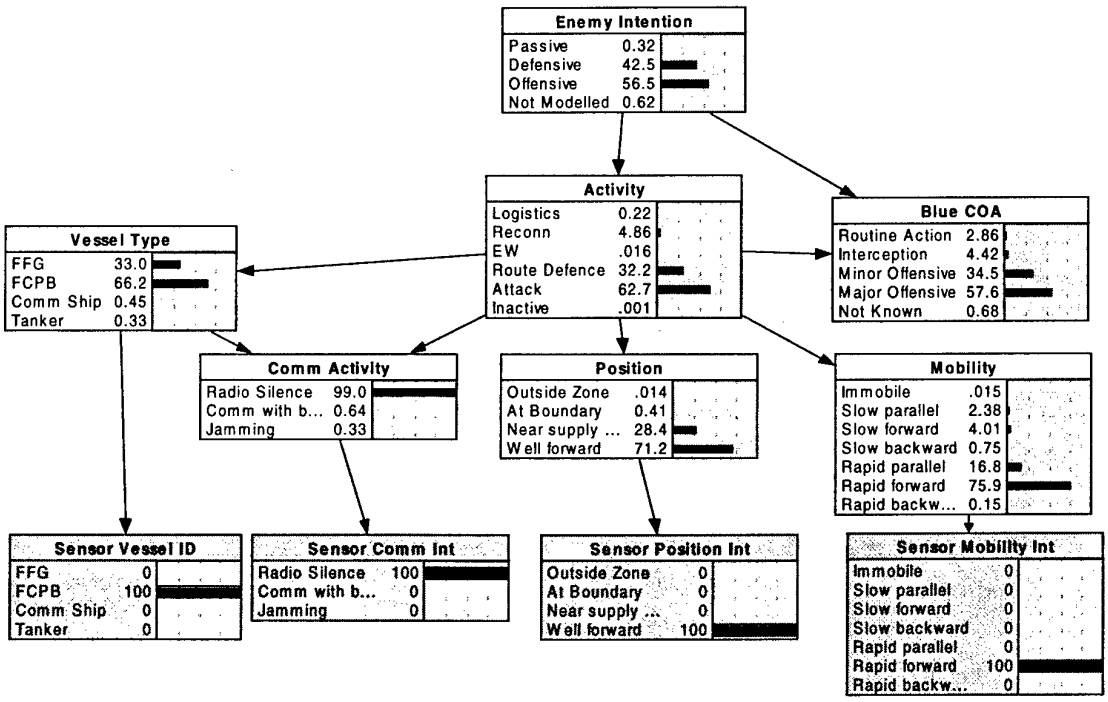

Figure D.4. The updated probability distribution P_2.

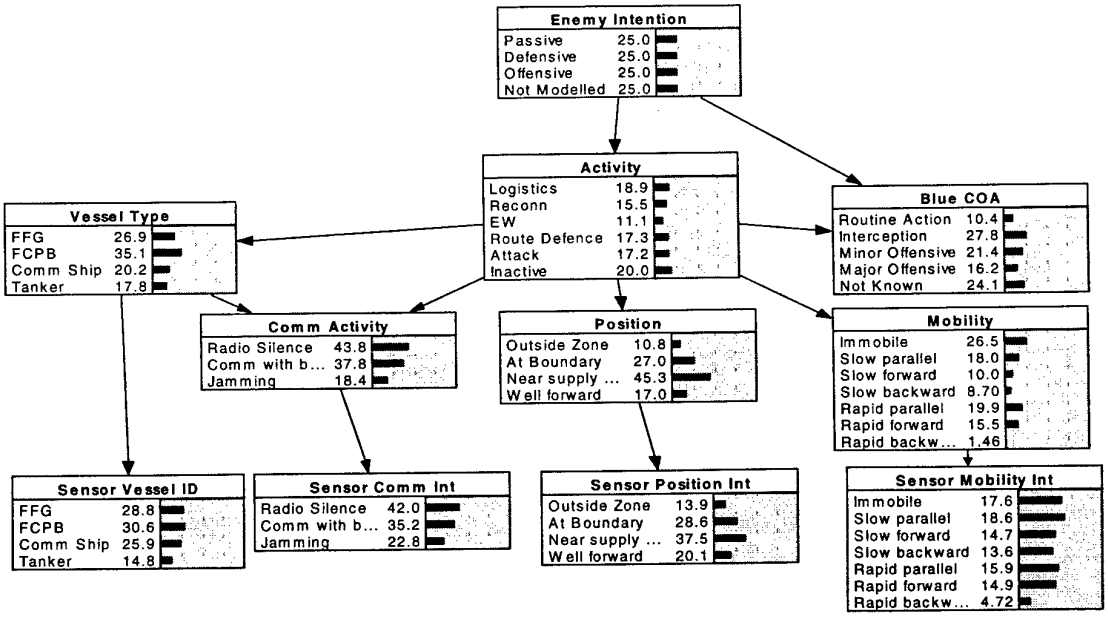

Figure D.5. Prior probability distribution corresponding to maximum entropy.

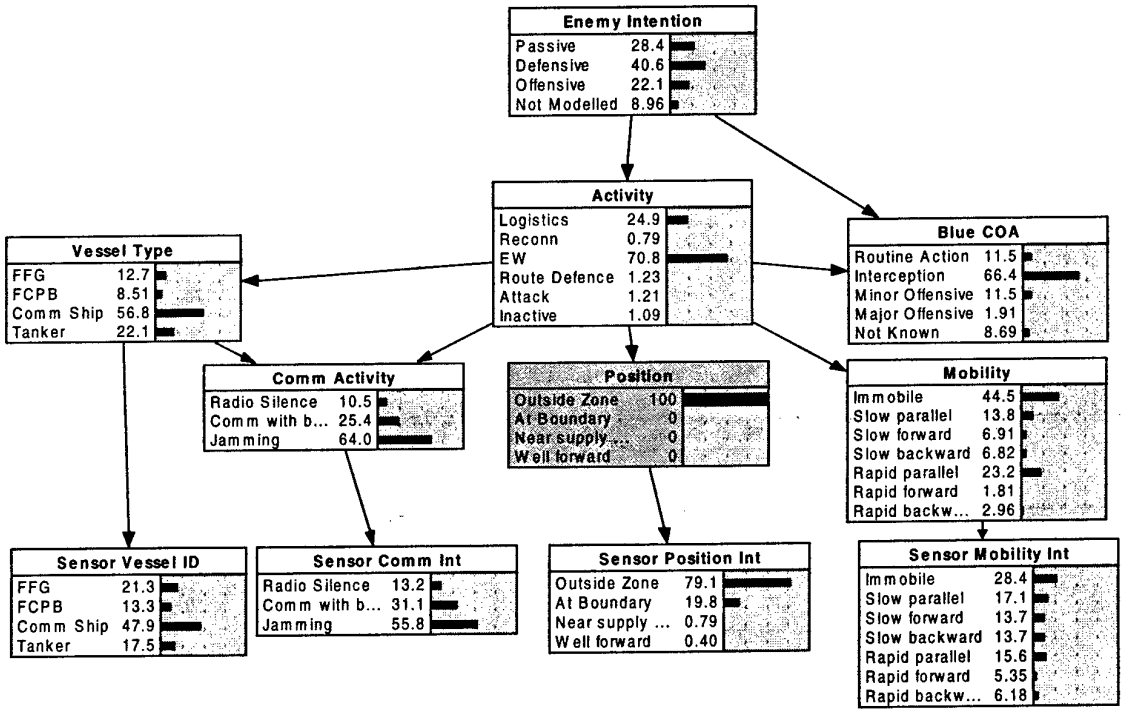

Figure D.6. Posterior probability distribution with *Position = Outside Zone*

DISTRIBUTION LIST

Representing Uncertainties Using Bayesian Networks

Balaram Das

AUSTRALIA

DEFENCE ORGANISATION

Task Sponsor

Director General C3I Development	1

S&T Program

Chief Defence Scientist FAS Science Policy AS Science Corporate Management	1shared copy
Director General Science Policy Development	1
Counsellor Defence Science, London	(Doc Data Sheet)
Counsellor Defence Science, Washington	(Doc Data Sheet)
Scientific Adviser to MRDC Thailand	(Doc Data Sheet)
Scientific Adviser Policy and Command	1
Navy Scientific Adviser	(Doc Data Sheet and distribution list only)
Scientific Adviser - Army	(Doc Data Sheet and distribution list only)
Air Force Scientific Adviser	1
Director Trials	1

Aeronautical and Maritime Research Laboratory

Director	1

Electronics and Surveillance Research Laboratory

Director	(Doc Data Sheet and distribution list only)
Chief of Information Technology Division	1
Research Leader CCIS	1
Research Leader ACC	(Doc Data Sheet)
Research Leader MCS	(Doc Data Sheet)
Research Leader C3	(Doc Data Sheet)
Research Leader Joint Systems	(Doc Data Sheet)
HSSA/ITD	1
HHSI/ITD	1
HIWS/ITD	(Doc Data Sheet)
HSSE/ITD	(Doc Data Sheet)
HC2AST/ITD	(Doc Data Sheet)
HTCS/ITD	(Doc Data Sheet)
HC3IOA/ITD	(Doc Data Sheet)

HIMF/ITD	(Doc Data Sheet)
HC3ISC/ITD	(Doc Data Sheet)
HDS/ITD	(Doc Data Sheet)
Author: Balaram Das, ITD	3
Mark Nelson, ITD	1
Rod Staker, ITD	1
J. Nicholson, LOD	1
D. Fogg, SSD	1
P. Berry, SSD	1
Publication and Publicity Officer ITD/ Executive Officer ITD	1

DSTO Library and Archives

Library Fishermens Bend	1
Library Maribyrnong	1
Library Salisbury	2
Australian Archives	1
Library, MOD, Pyrmont	(Doc Data sheet only)
US Defense Technical Information Center	2
UK Defence Research Information Centre	2
Canada Defence Scientific Information Service	1
NZ Defence Information Centre	1
National Library of Australia	1

Capability Systems Staff

Director General Maritime Development	(Doc Data Sheet only)
Director General C3I Development	(Doc Data Sheet only)
Director General Aerospace Development	(Doc Data Sheet only)

Navy

SO (Science), Director of Naval Warfare, Maritime Headquarters Annex, Garden Island, NSW 2000.	(Doc Data Sheet only)

Army

ABCA Standardisation Officer, Puckapunyal	4
SO (Science), DJFHQ(L), MILPO Enoggera, Queensland 4051	(Doc Data Sheet only)
NAPOC QWG Engineer NBCD c/- DENGRS-A, HQ Engineer Centre Liverpool Military Area, NSW 2174	(Doc Data Sheet only)

Intelligence Program

DGSTA Defence Intelligence Organisation	1
Manager DIO Information Centre	1

Corporate Support Program

OIC TRS, Defence Regional Library, Canberra	1

UNIVERSITIES AND COLLEGES

Australian Defence Force Academy	
Library	1
Head of Aerospace and Mechanical Engineering	1
Serials Section (M list), Deakin University Library, Geelong, 3217	1
Senior Librarian, Hargrave Library, Monash University	1
Librarian, Flinders University	1

OTHER ORGANISATIONS

NASA (Canberra)	1
AGPS	1
State Library of South Australia	1
Parliamentary Library, South Australia	1

OUTSIDE AUSTRALIA

ABSTRACTING AND INFORMATION ORGANISATIONS

Library, Chemical Abstracts Reference Service	1
Engineering Societies Library, US	1
Materials Information, Cambridge Scientific Abstracts, US	1
Documents Librarian, The Center for Research Libraries, US	1

INFORMATION EXCHANGE AGREEMENT PARTNERS

Acquisitions Unit, Science Reference and Information Service, UK	1
Library - Exchange Desk, National Institute of Standards and Technology, US	1
National Aerospace Laboratory, Japan	1
National Aerospace Laboratory, Netherlands	1

SPARES	5
Total number of copies:	**61**

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Representing Uncertainties Using Bayesian Networks

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHOR(S)

Balaram Das

5. CORPORATE AUTHOR

Electronics and Surveillance Research Laboratory
PO Box 1500
Salisbury SA 5108 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-0918	AR-011-177	Technical Report	December 1999

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
N8316/18/3	JNT99/138	DGC3ID	66	51

13. DOWNGRADING/DELIMITING INSTRUCTIONS

14. RELEASE AUTHORITY

Chief, Information Technology Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, SALISBURY, SA 5108

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CASUAL ANNOUNCEMENT Yes

18. DEFTEST DESCRIPTORS

Bayes Theorem, Bayesian Networks, Information Theory, Modelling, Situation Awareness, Uncertainty

19. ABSTRACT

This report demonstrates the application of Bayesian networks for modelling and reasoning about uncertainties. A scenario for naval anti-surface warfare is constructed and Bayesian networks are used to represent and update uncertainties encountered in the process of 'situation assessment'. Concepts from information theory are used to provide a measure of uncertainty and understand its flow in a Bayesian network. This in turn yields analytical methods to formulate various effectiveness measures.